MONTE CARLO METHODS: APPLICATION TO HYDROGEN GAS AND HARD SPHERES

BY

MARK DOUGLAS DEWING

B.S., Michigan Technological University, 1993
M.S., University of Illinois at Urbana-Champaign, 1995

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2001

Urbana, Illinois

MONTE CARLO METHODS: APPLICATION TO HYDROGEN GAS AND HARD SPHERES

Mark Douglas Dewing, Ph.D.
Department of Physics
University of Illinois at Urbana-Champaign, 2001
David M. Ceperley, Advisor

Quantum Monte Carlo (QMC) methods are among the most accurate for computing ground state properties of quantum systems. The two major types of QMC we use are Variational Monte Carlo (VMC), which evaluates integrals arising from the variational principle, and Diffusion Monte Carlo (DMC), which stochastically projects to the ground state from a trial wave function. These methods are applied to a system of boson hard spheres to get exact, infinite system size results for the ground state at several densities.

The kinds of problems that can be simulated with Monte Carlo methods are expanded through the development of new algorithms for combining a QMC simulation with a classical Monte Carlo simulation, which we call Coupled Electronic-Ionic Monte Carlo (CEIMC). The new CEIMC method is applied to a system of molecular hydrogen at temperatures ranging from 2800K to 4500K and densities from 0.25 to 0.46 g/cm^3.

VMC requires optimizing a parameterized wave function to find the minimum energy. We examine several techniques for optimizing VMC wave functions, focusing on the ability to optimize parameters appearing in the Slater determinant.

Classical Monte Carlo simulations use an empirical interatomic potential to compute equilibrium properties of various states of matter. The CEIMC method replaces the empirical potential with a QMC calculation of the electronic energy. This is similar in spirit to the Car-Parrinello technique, which uses Density Functional Theory for the electrons and molecular dynamics for the nuclei.

The challenges in constructing an efficient CEIMC simulation center mostly around the noisy results generated from the QMC computations of the electronic energy. We introduce two complementary techniques, one for tolerating the noise and the other for reducing it. The penalty method modifies the Metropolis acceptance ratio to tolerate noise without introducing a bias in the simulation of the nuclei. For reducing the noise, we introduce the two-sided energy difference method, which uses correlated sampling to compute the energy change associated with a trial move of the nuclear coordinates. Unlike the standard reweighting method, it remains stable as the energy difference increases.

Acknowledgments

First I would like to thank my advisor, David Ceperley, for supporting me in this research, for teaching these Monte Carlo methods to me, and for being available and patient when answering questions.

I would also like to thank the graduate students and postdocs in the group who have helped me and provided useful and interesting discussions. And special thanks to Tadashi Ogitsu for the use of his DFT code.

I am grateful to my parents for their support during my college and graduate school pursuits. I enjoyed the refreshing summer visits to their farm. Thanks to my brother Luke, whose living in Colorado was convenient for ski trips.

During my time here, I have benefited greatly from friendships and relationships with people in Graduate Intervarsity Christian Fellowship, Grace Community Church, and Illini Life Christian Fellowship. They have given me a great deal of strength and encouragement when I needed it.

My work was supported by the computational facilities at NCSA, by a Graduate Research Trainee fellowship NSF Grant No. DGE93-54978, and by NSF Grant No. DMR 98-02373.

And finally, apologies to my cat, Lucy, for not giving her enough attention while finishing this work.

Contents

List of Tables

List of Figures

Guide to Notation

CEIMC	Coupled Electronic-Ionic Monte Carlo
DMC	Diffusion Monte Carlo
DFT	Density Functional Theory
GBRW	Gradient Biased Random Walk
LDA	Local Density Approximation (in Density Functional Theory)
PIMC	Path Integral Monte Carlo
QMC	Quantum Monte Carlo
SGA	Stochastic Gradient Approximation
VMC	Variational Monte Carlo
α	General variational parameter.
a	Instantaneous acceptance probability in penalty method.
a_0	Bohr radius, unit of length in atomic units. $1\ a_0 = 5.29 \times 10^{-9}$ m.
$A(s \to s')$	Acceptance probability in Metropolis method.
β	Inverse temperature, $\frac{1}{k_B T}$.
δ	QMC estimate of an energy difference.
Δ	Exact energy difference. Also the sampling box size in the Metropolis algorithm.
d	Bond length.

D	Slater determinant.
ζ	A variational parameter in H_2 orbitals.
f	A trial wave function (also denoted ψ_T). Also the relative noise parameter.
η	Additional noise rejection ratio.
E_L	The local energy of a trial wave function.
θ_v	Vibrational temperature.
h	Step size parameter in SGA and GBRW.
H	A many-body Hamiltonian.
Ha	Hartree, unit of energy in atomic units. 1 Ha = 27.21 eV.
$G(R \rightarrow R', \tau)$	Green's function propagator in DMC.
k_B	Boltzmann factor.
K	Kinetic energy.
λ	$\hbar^2/2m$, where m is the particle's mass.
n_0	Condensate fraction.
N	The number of particles in a system.
$\pi(s)$	Probability distribution to be sampled in a Markov process.
P	Pressure, or a probability distribution in the two-sided method.
$P(s \rightarrow s')$	Transition probability in a Markov chain.
Q	Normalization integral .
ρ	Density.
$\rho_1(r)$	Single particle density matrix.
r_s	$\left(\frac{3}{4\pi n}\right)^{1/3}$, where n is the electron number density.
r_{ij}	The separation between particle j and particle i.

R	The coordinates of all the particles in a many-body system.
σ	Noise level (variance or standard error). Also the hard sphere diameter.
s	State in configuration space.
τ	DMC time step .
T	Temperature.
$T(s \rightarrow s')$	Sampling distribution in Metropolis method.
$u(r)$	Jastrow factor in a many-body wave function.
V	Volume of the simulation cell, or potential energy.
V	Alternate notation for potential energy (for formulas that have both potential energy and volume).
w	Weight factor in correlated sampling methods.
w_l	Width of H_2 orbital.
ψ_T	A trial wave function (also denoted f).
ϕ_0	The exact many body ground state wave function.
ϕ	A single particle orbital.

Chapter 1

Introduction

The first computer simulations of a condensed matter system used the simplest potential, the hard sphere (**?**). As computers and simulations progressed, more sophisticated and realistic potentials came into use. These potentials are parameterized and then fit to reproduce various experimental quantities. Both Molecular Dynamics (MD) and Monte Carlo (MC) methods are used to generate ensemble averages of many-particle systems.

These potentials originate from the microscopic structure of matter, described in terms of electrons, nuclei, and the Schrödinger equation. But the many-body Schrödinger equation is too complicated to solve directly, so some approximations are needed. The one electron approximation is a successful approach, where a single electron interacts with an external potential (ie, the nuclei) and with a mean field generated by all the other electrons. This is done by Hartree-Fock (HF) or with Density Functional Theory (DFT) (**?**). DFT is in principle exact, but contains an unknown exchange and correlation functional that must be approximated. The most common one is the Local Density Approximation (LDA).

These first principles calculations are used in fitting the potentials, which are then used in an MC or MD computation. But the problem of transferability still remains. Empirical potentials are only valid in situations for which they have been designed and fitted.

In 1985, Car and Parrinello introduced their method, which replaced the empirical potential with a DFT calculation done 'on the fly' (**?**). They did a molecular dynamics simulation of the nuclei of liquid silicon and then computed the density functional energy of the electrons at every MD step. To improve the efficiency of the computation of the DFT energy, they introduced a new iterative method for solving the DFT equations. It has been a very successful method, with the original paper being cited over 2300 times since its publication.

Previously, the DFT equations had been solved by eigenvalue methods. But eigenvalue problems can also be regarded as optimization problems, where an energy functional is

minimized. Car and Parrinello used an idea similar to simulated annealing, but they used molecular dynamics to move through parameter space, rather than Monte Carlo. This had the effect of making the equations of motion similar between the electronic problem and the nuclear problem, with the only difference being in the relative masses. Since the electronic problem was not real electron dynamics, the electron mass does not correspond to any physical quantity, and is only a parameter controlling the convergence of the electronic part of the simulation. Since then, other iterative methods have been introduced, usually based on the Conjugate-Gradient method (**?**).

A brief review of applications of the Car-Parrinello method to liquid problems is given by **?**). This review also mentions that LDA and some other functionals are not good enough to accurately simulate water (there are improved functionals that are acceptable). Another review of molecular dynamics by **?**) includes material on treating the nuclei classically and also using path integrals to treat the nuclei quantum mechanically, done by **?**).

Quantum Monte Carlo (QMC) methods have developed as another means for accurately solving the many body Schrödinger equation (**???**). The success of QMC lies partly in the fact these methods explicitly include correlation among the electrons, which can not be done directly with the one electron methods. Particularly with the Local Density Approximation (LDA), DFT has known difficulties in handling electron correlation (**?**).

In the spirit of the Car-Parrinello method, we integrate a Classical Monte Carlo simulation of the nuclei with a QMC simulation for the electrons. This we call Coupled Electronic-Ionic Monte Carlo (CEIMC). There are some challenges in constructing an efficient method.

The first problem we encounter is that the results of a QMC simulation are noisy. The QMC energy has some uncertainty associated with it, and it could bias the classical part of the simulation. We could run the QMC simulation until the noise is negligible, but that is very time-consuming. A better way is use the penalty method, which modifies the usual MC formulas to be tolerant of noise.

The electrons are assumed to be in their ground state, both in the Car-Parrinello method and in our CEIMC method. There are two internal effects that could excite the electrons - coupling to nuclear motion and thermal excitations. In the first case, we make the Born-Oppenheimer approximation, where the nuclei are so much more massive than the electrons that the electrons are assumed to respond to nuclear motion instantaneously, and so stay in their ground state. We neglect any occupation of excited states of the electrons due to coupling to nuclear motion.

In the case of thermal excitation, let us examine several relevant energy scales. If we consider a gas of degenerate electrons at a density of $n = 0.0298$ electrons per cubic Bohr

(i.e. $r_s = \left(\frac{3}{4\pi n}\right)^{1/3} = 2.0$), the Fermi temperature is about 140,000K. The gap between the ground state and the first excited state of a hydrogen molecule at equilibrium bond distance is about 124,000K. As long as our temperatures are well below this (and they are), and we are not at too high pressures (pressure decreases the gap), the thermal occupation of excited states can be neglected.

Hydrogen is the most abundant element in the universe, making an understanding of its properties important, particularly for astrophysical applications. Models of the interiors of the gas giant planets depends on a knowledge of the equation of state of hydrogen (**??**). Hydrogen is also the simplest element, but it still displays remarkable variety in its properties and phase diagram. It has several solid phases at low temperature, and the crystal structure of one of them (phase III) is not fully known yet. At high temperature and pressure the fluid becomes metallic, but the exact nature of the transition is not known.

Computer simulation can also be used to obtain results on model systems. We will examine the hard sphere Bose gas, a simple and important model. For this model, all the approximations we make are controllable, and we will look at how to deal with those approximations and obtain exact results for this model.

1.1 Thesis Overview

Chapter 2 is an introduction to the basic classical and quantum Monte Carlo techniques we will be using. Chapter 3 presents an improved QMC method for computing the energy difference between two systems. Chapter 4 is an examination of parameter optimization, which is essential in VMC. We present various methods for minimizing the energy, and give some comparisons between them.

Successful CEIMC simulations are based on the penalty method for tolerating noise in the Metropolis method, which is detailed in Chapter 5. Some additional details are discussed, and an example of CEIMC applied to a single H_2 molecule is given. The results of computations of the ground state energy of the boson hard sphere model are presented in Chapter 6. In Chapter 7, the CEIMC simulation method is applied to fluid molecular hydrogen. We present data for a few state points and perform some analysis of the simulation itself.

Chapter 2

Monte Carlo Methods

Monte Carlo integration methods are very useful for evaluating the basic integrals of statistical and quantum physics. In a system with N_p particles, these integrals have the form

$$\langle O \rangle = \frac{\int dR\, \pi(R) O(R)}{\int dR\, \pi(R)} \tag{2.1}$$

where R is a $3N_p$ dimensional vector, $\pi(R)$ is a probability distribution, and $O(R)$ is the observable or quantity of interest. These integrals have two important characteristics: high dimensionality and the integrands are sharply peaked - only small parts of phase space contribute significantly to the integral.

The high dimensionality makes a grid based scheme impractical in two ways. First, suppose we have a 300 dimensional integral (100 particle simulation), and want 10 grid points in each dimension. Even this crude integration requires function evaluations at 10^{300} grid points! Second, consider the trapezoidal rule (as a concrete example) in d dimensions. The error using N samples will go as $O(N^{-2/d})$. As we will show, the error in Monte Carlo integration goes as $O(N^{-1/2})$. The Monte Carlo error is independent of the dimensionality whereas the grid based method depends on it strongly. For these high dimensionality problems, Monte Carlo is only practical choice.

2.1 Basic Monte Carlo Integration

Consider an integral of the form

$$I = \int_0^1 f(x) dx. \tag{2.2}$$

To evaluate by Monte Carlo, compute $f(x)$ at N points sampled uniformly from $[0,1]$. An approximation to I is given by

$$I \approx \bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(x_i) \tag{2.3}$$

The estimate of the statistical error in $\bar{f}$ will be

$$\sigma_I = \sigma_f / \sqrt{N} \tag{2.4}$$

where σ_f^2 is the variance, and is given by

$$\sigma_f^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(f(x_i) - \bar{f} \right)^2 \tag{2.5}$$

Thus the error goes as $O(N^{-1/2})$.

The error bounds can be improved by sampling more points, or by reducing the variance, σ_f^2. The latter can be accomplished with importance sampling. Consider some probability, $P(x)$, that is an approximation to $f(x)$. Write Eq. (2.2) as

$$I = \int_0^1 P(x) \frac{f(x)}{P(x)} dx. \tag{2.6}$$

The estimate of I is obtained by sampling N points from $P(x)$ and computing

$$I \approx \sum_{i=1}^{N} \frac{f(x_i)}{P(x_i)} \tag{2.7}$$

If P is a good approximation to f, then the variance of the sum in Eq. (2.7) is much less than the variance of the sum in Eq. (2.3).

The fact that the integrands of interest are sharply peaked, as mentioned previously, makes importance sampling a necessity. The most useful type of importance sampling for these problems is the Metropolis method.

2.2 Metropolis Sampling

The Metropolis method (**?**) uses a Markov process to generate samples from a normalized probability distribution, $\pi(R) / \int dR \; \pi(R)$. These samples are then used to estimate Eq. (2.1) by

$$\bar{O} = \frac{1}{N} \sum_i O(R_i) \tag{2.8}$$

For generality in the following section, we will denote the state of the simulation by s.

A Markov process takes a transition probability between states, $P(s \to s')$, and constructs a series of state points $s_1, s_2, \ldots$ (called a chain). An important characteristic of a Markov process is that the choice of the next state point in the chain depends only on the current state point, not any previous state points.

The Metropolis method constructs a transition probability such that generated state points are sampled from the desired distribution. For this to work, the transition probability must satisfy ergodicity. This means the Markov chain must eventually be able to reach any state in the system. A sufficient condition for satisfying ergodicity is detailed balance,

$$\pi(s)P(s \to s') = \pi(s')P(s' \to s). \tag{2.9}$$

The generalized Metropolis method breaks the transition probability into the product of two pieces - an *a priori* sampling distribution $T(s \to s')$ and an acceptance probability $A(s \to s')$. The Metropolis choice for the acceptance probability is

$$A(s \to s') = \min\left[1, \frac{\pi(s')T(s' \to s)}{\pi(s)T(s \to s')}\right] \tag{2.10}$$

The procedure is to sample a trial state, s', according to $T(s \to s')$ and evaluate Eq. (2.10). The acceptance probability is compared with a uniform random number on $[0,1]$. If A is greater than the random number, the move is accepted, s' becomes the new s and is used in the average in Eq (2.8). Otherwise the move is rejected, s is not changed, and is reused in the average.

The original Metropolis procedure chooses a trial position uniformly inside a box centered around the current point,

$$s' = s + y \tag{2.11}$$

where y is a uniform random number on $[-\Delta/2, \Delta/2]$, with Δ being an adjustable parameter. In this case, T is uniform and will cancel out of Eq. (2.10).

An important measure of the Metropolis procedure is the acceptance ratio - the ratio of accepted moves to the number of trial moves. It can be adjusted by the choice of Δ. If the acceptance ratio is too small, state space will be explored very slowly because very few moves are accepted. If the acceptance ratio is high, it is likely that the trial moves are too small and once again, diffusion through state space will be very slow. Balancing these considerations leads to the standard rule of thumb that the optimal acceptance ratio is around 50%.

A better consideration is maximization of the efficiency,

$$\xi = \frac{1}{\sigma^2 T} \tag{2.12}$$

where T is the computer time taken to get an error estimate of σ.

2.3 Classical Monte Carlo

The probability distribution we wish to sample is the Boltzmann distribution

$$\pi(s) \propto \exp[-V(s)/kT]. \tag{2.13}$$

The first simulations of this type were done with the hard sphere potential (**??**). Later simulations used Lennard-Jones potentials, and then other types of empirical potentials.

The Metropolis procedure samples only the normalized $\pi(s)$. Averages over this distribution are readily computed, but quantities that depend on the value of the normalization are difficult to compute. In classical systems, this includes quantities such the entropy and the free energy. There are techniques, however, for computing the free energy difference between two systems.

2.4 Variational Monte Carlo

Variational Monte Carlo (VMC) is based on evaluating the integral that arises from the variational principle. The variational principle states that the energy from applying the Hamiltonian to a trial wave function must be greater than or equal to the exact ground state energy. Typically the wave function is parameterized and then optimized with respect to those parameters to find the minimum energy (or minimum variance of the energy). Monte Carlo is needed because the wave function contains explicit two (or higher) particle correlations and this results in a non-factoring high dimensional integral.

The energy is written as

$$E = \frac{\int dR \psi_T(R) H \psi_T(R)}{\int dR \psi_T(R)^2} = \frac{\int dR |\psi_T(R)|^2 E_L(R)}{\int dR |\psi_T(R)|^2} \tag{2.14}$$

where $E_L = \frac{H\psi_T}{\psi_T}$, and is called the local energy. Other diagonal matrix elements can be evaluated in a similar fashion. Off diagonal elements can also be evaluated, but with more effort. **?**) introduced the use of Metropolis sampling for evaluating this integral.

A typical form of the variational wave function is a Jastrow factor (two body correlations) multiplied by two Slater determinants of one body orbitals.

$$\psi_T = \exp\left[-\sum_{i<j} u(r_{ij})\right] \mathrm{Det}\left(S^{\uparrow}\right) \mathrm{Det}\left(S^{\downarrow}\right) \tag{2.15}$$

The Jastrow factor, u, will contain electron-nucleus and electron-electron correlations. Appendix B has details on the derivatives that enter into the local energy.

As two electrons or an electron and a nucleus get close, there is a singularity in the Coulomb potential. That singularity needs to be canceled by kinetic energy terms in the wave function. This requirement is known as the cusp condition. Details are given in Appendix C.

Techniques for the efficient handling of the determinants were developed by **?**). The VMC algorithm is implemented so that only single electron trial moves are proposed. This causes a change in only one column of the Slater matrix. The new determinant and its derivatives can be computed in $O(N)$ operations, given the inverse of the old Slater matrix. This inverse is computed once at the beginning of the simulation and then updated whenever a trial move is accepted. The update takes $O(N^2)$ operations. By comparison, computing the determinant directly takes $O(N^3)$ operations. This technique creates a situation where there is more work done for an acceptance than for a rejection. A lower acceptance ratio will be faster, since fewer updates need to be performed. Details of the updating procedure and some other properties of determinants are given in Appendix A.

Optimization of the parameters in the wave function is a large topic, so we will defer the discussion until a later chapter. However, we will make one observation here. If ψ_T is an eigenstate, the local energy becomes constant and any MC estimate for the energy will have zero variance. This zero-variance principle allows searching for optimum parameters by minimizing the variance rather than minimizing the energy. In principle this is true for any eigenstate, not just the ground state.

2.4.1 Two Level Sampling

A multilevel sampling approach can be used to increase the efficiency of VMC (**?**). Multilevel sampling has been used extensively in path integral Monte Carlo (**?**). The general idea is to use a coarse approximation to the desired probability function for an initial accept/reject step. If it is accepted at this first level, a more accurate approximation is used, and another accept/reject step is made. This continues until the move is rejected or until the most detailed level has been reached. This method increases the speed of the calculation because the entire probability function need not be computed every time.

Consider splitting the wave function into two factors - the single body part (D) and the two body part (e^{-U}). Treat the single body part as the first level, and the whole wave function as the second level.

First, a trial move, R', is proposed and accepted with probability

$$A_1 = \min\left[1, \frac{D^2(R')}{D^2(R)}\right] \tag{2.16}$$

If it is accepted at this stage, the two body part is computed and the trial move is accepted with probability

$$A_2 = \min\left[1, \frac{\exp[-2U(R')]}{\exp[-2U(R)]}\right] \tag{2.17}$$

It can be verified by substitution that this satisfies detailed balance in Eq. (2.9). After an acceptance at this second level, the inverse Slater matrices are updated, as described previously.

We compared the efficiency between the standard sampling method and the two level sampling method on two test systems: a single Li_2 molecule in free space, and a collection of 32 H_2 molecules in a periodic box of side 19.344 a.u. (r_s = 3.0). The wave functions are taken from **?**), and will be described in Section 2.7.

The step size, Δ, is the obvious parameter to adjust in maximizing the efficiency. But we can also vary the number of steps between computations of E_L. The Metropolis method produces correlated state points (see more on serial correlations in Section 2.6), so successive samples of E_L do not contain much new information. In these tests we sampled E_L every five steps.

Results for the different sampling methods with Li_2 are shown in Tables 2.1 and 2.2. The Determinant Time and Jastrow Time columns include only the time needed for computing the wave function ratio in the Metropolis method, and not the time for computing the local energy. The total time column does include the time for computing the local energy. The efficiency is also shown on the left in Figure 2.1.

For the two level method with Li_2, the second level acceptance ratio is quite high, indicating the single body part is a good approximation to the whole wave function.

Results for the collection of H_2 molecules are given in Tables 2.3 and 2.4. The efficiency is also shown on the right graph in Figure 2.1.

Comparing the maximum efficiency for each sampling method, two level sampling is 39% more efficient than standard sampling for Li_2, and 72% more efficient for the collection of H_2's.

2.5 Diffusion Monte Carlo

Diffusion Monte Carlo (DMC) is a method for computing the ground state wave function. It typically takes an order of magnitude more computing time than VMC, and is most efficient when used in conjunction with a good VMC trial function.

Formally, DMC and related methods work by converting the differential form of the Schrödinger equation into an integral equation and solving that integral equation by stochastic methods. From another point of view, the Schrödinger equation in imaginary time and

Table 2.1: Timings for Li_2 molecule using the standard sampling method. All times are in seconds on an SGI Origin 2000.

Δ	Acceptance Ratio	Determinant Time	Jastrow Time	Total Time	ξ
1.0	0.610	48.3	340	516	1190
1.5	0.491	48.1	340	508	1680
2.0	0.407	48.2	340	503	1460
2.5	0.349	48.2	339	499	1070
3.0	0.307	48.2	339	496	800

Table 2.2: Timings for Li_2 molecule using the two level sampling method. All times are in seconds on an SGI Origin 2000.

Δ	First Level Acc. Ratio	Second Level Acc. Ratio	Total Acc. Ratio	Time	ξ
1.0	0.674	0.899	0.606	400	1580
1.5	0.543	0.894	0.485	347	2430
2.0	0.447	0.897	0.401	304	2340
2.5	0.379	0.902	0.342	276	1910
3.0	0.331	0.906	0.300	256	1400

Table 2.3: Timings for the system of 32 H_2 molecules in a periodic box using the standard sampling method. All times are in seconds on a Sun Ultra 5.

Δ	Acceptance Ratio	Determinant Time	Jastrow Time	Total Time	ξ
2.0	0.606	167	1089	2015	0.61
3.0	0.455	167	1085	1891	1.22
4.0	0.338	166	1084	1794	1.23
5.0	0.250	166	1080	1722	1.06
6.0	0.185	164	1080	1668	1.02
7.0	0.139	162	1084	1629	0.76

Table 2.4: Timings for a system of 32 H_2 molecules in a periodic box using the two level sampling method. All times are in seconds on a Sun Ultra 5.

Δ	First Level Acc. Ratio	Second Level Acc. Ratio	Total Acc. Ratio	Total Time	ξ
2.0	0.740	0.795	0.589	1804	0.59
3.0	0.598	0.728	0.436	1421	1.77
4.0	0.468	0.681	0.319	1185	2.11
5.0	0.357	0.649	0.232	994	1.55
6.0	0.370	0.627	0.169	849	1.87
7.0	0.204	0.609	0.124	740	1.46

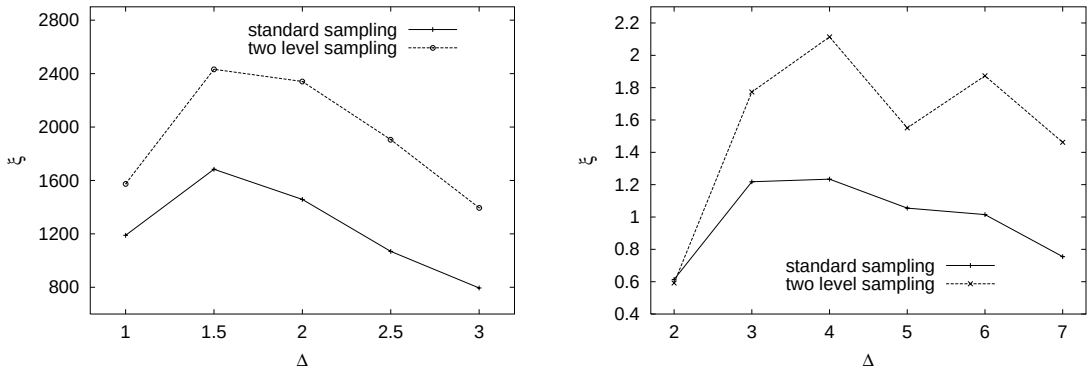


Figure 2.1: Efficiency of VMC. The graph on the left is for Li_2. The graph on the right is for 32 H_2 molecules.

the diffusion equation are very similar, enabling one to use a random process to solve the imaginary time Schrödinger equation. This similarity was recognized early and was proposed as a computational scheme in the early days of computing (**?**). Unfortunately, without importance sampling, it is very inefficient computationally.

The ground state wave function can be obtained by the projection

$$\phi_0 = \lim_{\tau\to\infty} e^{-\tau(H-E_0)}\psi_T \tag{2.18}$$

where E_0 is the ground state energy. This can be seen by expanding ψ_T in energy eigenstates,

$$\begin{aligned} e^{-\tau(H-E_0)}\psi_T &= e^{-\tau(H-E_0)}\sum_i \phi_i \\ &= \sum_n e^{-\tau(E_n-E_0)}\phi_n. \end{aligned} \tag{2.19}$$

At large τ, the contribution from the excited states will decay exponentially, and only the ground state will remain. To make a practical computation method, we write the projection in the position basis as

$$\psi(R',t+\tau) = \int dR\psi(R,t)G(R\to R';\tau) \tag{2.20}$$

where $G = \langle R'|e^{-\tau H}|R\rangle$ and $f(R,t)$ is the wave function after some time t. This equation is iterated to get to the large time limit. The fully interacting, many-body Green's function is too hard to compute, so the various methods differ in how they approximate the full projector. In particular, DMC makes a short time approximation, and the resulting pieces have natural interpretations in terms of a diffusion process with branching. The name Projector Monte Carlo or Green's Function Monte Carlo is often applied to these methods. Perhaps unfortunately, the name Green's Function Monte Carlo (GFMC) is also applied to a specific technique that uses a spatial domain decomposition for the Green's function.

For a more detailed presentation of DMC, with importance sampling, we mostly follow **?**). We start with the Schrödinger equation in imaginary time

$$-\frac{\partial\phi(R,t)}{\partial t} = \left[-\lambda\nabla^2 + V(R) - E_T\right]\phi(R,t) \tag{2.21}$$

where $\lambda = \hbar^2/2m$. Importance sampling is added by multiplying Eq. (2.21) by a known trial function ψ_T. The result, written in terms of the "mixed distribution" $f(R,t) = \phi(R,t)\psi_T(R)$, is

$$-\frac{\partial f(R,t)}{\partial t} = -\lambda\nabla^2 f + (E_L(R) - E_T)f + \lambda\nabla\cdot(fF_Q(R)) \tag{2.22}$$

where E_L is the local energy, as in VMC, and $F_Q = 2\nabla\psi_T/\psi_T$ (called the quantum force).

Once again, the solution for f in terms of a Green's function is

$$f(R',t+\tau) = \int dR f(R,t) G(R \to R';\tau) \tag{2.23}$$

For sufficiently short times, we can ignore the non-commutivity of the kinetic and potential terms in the Hamiltonian, $e^{-\tau H} \approx e^{-\tau T} e^{\tau V}$. The explicit form for the short time Green's function in the position basis is

$$\begin{aligned} G(R \to R';\tau) &= (4\pi\lambda\tau)^{-3N/2} G_{\text{branch}}(R \to R';\tau) G_{\text{drift}}(R \to R';\tau) & (2.24)\\ G_{\text{branch}}(R \to R';\tau) &= \exp\left[-\tau\{\bar{E} - E_T\}\right] & (2.25)\\ G_{\text{drift}}(R \to R';\tau) &= \exp\left[-\left\{R' - R - \lambda\tau F_Q(R)\right\}^2 / 4\lambda\tau\right] & (2.26) \end{aligned}$$

where $\bar{E} = \left[E_L(R) + E_L(R')\right]/2$.

The algorithm is started by generating a collection of configurations ("walkers"), usually sampled from ψ_T. Equation (2.23) proceeds by applying a drifting random walk to each particle. The new position of the ith particle is given by

$$\vec{r}_i' = \vec{r}_i + \lambda\tau\vec{F}_Q(R) + \sqrt{2\lambda\tau}\,\chi \tag{2.27}$$

where χ is a normally distributed random variable with zero mean and unit variance. In a simple interpretation of Eq. (2.23), this would always be the new position. But consider the case if ψ_T becomes the true ground state, ϕ_0. The branching term is then constant and the algorithm becomes similar to VMC. In this case we want to sample the correct distribution for any τ. This is done by adding a Metropolis rejection step, where the trial move is accepted with probability

$$A = \min\left[1, \frac{\psi_T(R')^2 G(R \to R',\tau)}{\psi_T(R)^2 G(R' \to R,\tau)}\right] \tag{2.28}$$

Each configuration is then weighted by G_{branch}. Because of rejections in the previous step, the time, τ, in Eq. (2.25) should be replaced by τ_{eff}, which is

$$\tau_{\text{eff}} = \frac{\langle r^2_{\text{accepted}}\rangle}{\langle r^2_{\text{total}}\rangle}\tau \tag{2.29}$$

where $\langle r^2_{\text{accepted}}\rangle$ is the mean square displacement of the accepted electron moves and $\langle r^2_{\text{total}}\rangle$ is the mean square displacement of all the proposed electron moves.

The weighting is done by adding or removing configurations from the collection (branching). This is done by computing the multiplicity $M = \text{int}(G_{\text{branch}} + y)$, where y is a random number on $[0,1]$. This multiplicity is the number of copies of this configuration that should be retained in the collection of walkers. If it is zero, the configuration is deleted from the

collection. If it is one, the configuration remains as is. If it is greater than one, additional copies of this configuration are added to the collection.

The number of walkers in the collection is kept roughly constant by adjusting E_T. In particular, the trial energy is adjusted according to

$$E_T = E_0 + \kappa \ln(P^*/P) \tag{2.30}$$

where E_0 is the best guess for the ground state energy, P is the current population, P^* is the desired population, and κ is a feedback parameter.

The energy is computed by averaging the local energy over the distribution of walkers. Once the transients have decayed away, subsequent steps are part of the ground state distribution. The program is then run for however long is necessary to gather statistics for the energy and other estimators.

There is a problem with DMC for estimating quantities other than the energy. The expectation value is not averaged over the ground state, but over the mixed distribution $\phi_0\psi_T$. This can be partly corrected by using the extrapolated estimator,

$$\langle\phi_0|A|\phi_0\rangle \approx 2\langle\phi_0|A|\psi_T\rangle - \langle\psi_T|A|\psi_T\rangle \tag{2.31}$$

.

Getting the correct estimator (also called a pure estimator) requires "forward walking", so named because the weight needed, ϕ_0/ψ_T, is related to the asymptotic number of children of each walker (**?**). This can be implemented by storing the value of the estimator and propagating it forward with the walker for a given number of steps (**?**).

2.5.1 Fermions

In all these methods, some quantity is treated as a probability, which requires that it be positive. In VMC this quantity is $|\psi_T|^2$, which is always positive. For DMC, we sample from $\psi_T\phi_0$, which can be negative if the fermion nodes of ψ_T are not the same as the nodes of ϕ_0. The simplest cure is to fix the nodes of the ground state to be the same as ψ_T. This is known as the fixed-node approximation. It is implemented in the DMC algorithm by rejecting moves that would change the sign of the determinant of ψ_T.

2.6 Statistical Errors

The formula for the variance given in Eq. (2.5) assumes that there are no serial correlations in the data. However, the Metropolis sampling method produces correlated data, which must be considered when estimating the statistical error.

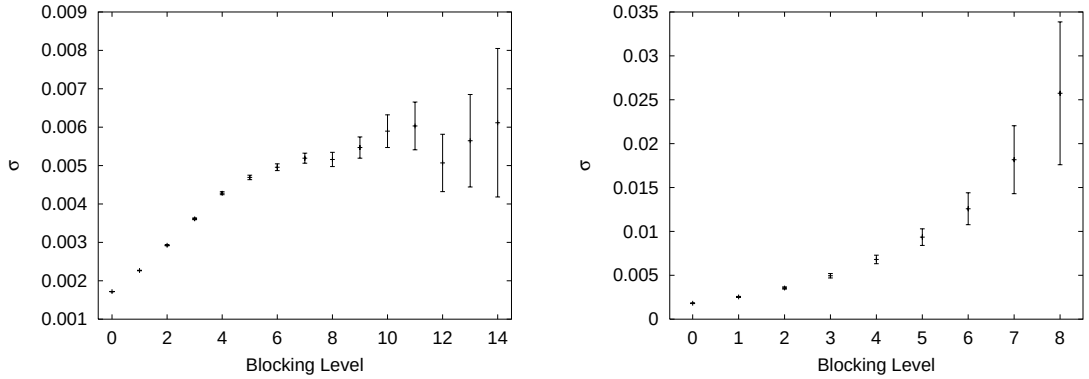


Figure 2.2: Examples of statistical data analysis using reblocking. The error in the graph on the left has converged, while the error in the graph on the right has not.

Correlations in data are quantified by the autocorrelation function, defined for some estimator, E, as

$$C(k) = \frac{1}{(N-k)} \sum_{i=1}^{N-k} (E_i - \bar{E})(E_{i+k} - \bar{E}) \tag{2.32}$$

The autocorrelation time, κ, is computed as

$$\kappa = 1 + \frac{2}{\sigma^2} \sum_{k=1}^{\text{cutoff}} C(k) \tag{2.33}$$

This sum tends to be quite noisy. As a heuristic strategy, we can approximate κ by the first place where the autocorrelation function drops below 10%. The true variance of the mean is the simple variance of the individual data points multiplied by κ.

Another way to estimate the true error is by reblocking (**??**). At the second level, take the average of every 2 data points. Now compute the variance of this set of data that has $N/2$ points. Continue this procedure recursively until the variance stops changing. **?**) gives a well-defined procedure for computing when that occurs. Figure 2.2 shows some example plots of error vs. reblocking level. On the left hand graph we see the expected plateau in the error estimate. On the right hand graph there is no plateau, indicating that there is not enough data to reliably estimate the error.

2.7 Wave Functions

For our studies of molecular hydrogen, we started with the wave function ψ_{III} from Reynolds (**?**). The Jastrow factors are

$$u_{ee} = -\sum_{i<j} \frac{a_e r_{ij}}{1 + b_{ee} r_{ij}}$$

$$u_{ne} = \sum_{i,\alpha} \frac{Z_\alpha a_n r_{i\alpha}}{1 + b_{en} r_{i\alpha}} \tag{2.34}$$

where Z is the nuclear charge and b is the variational parameter. The cusp conditions are satisfied by setting $a_n = 1$ and $a_e = 1/2$. As noted in Appendix C, having the correct cusp condition for parallel spins does not affect the energy much, so the same value for a_e is used for parallel and antiparallel electron spins. The $b's$ from the two types of Jastrow factors are folded into a single parameter, $\beta = a/b^2$.

The orbitals are floating Gaussians, with the form

$$\phi_l(r) = \exp\left[\frac{-(r - c_l)^2}{w_l^2}\right] \tag{2.35}$$

where c_l is the center of orbital, and w_l is a free parameter. In molecular hydrogen, c_l will be fixed at the bond center.

The orbitals can be generalized to be anisotropic,

$$\phi_l(r) = \exp\left[-(\mathbf{r} - \mathbf{c}_l)^T \cdot \mathbf{R}^T \Gamma \mathbf{R} \cdot (\mathbf{r} - \mathbf{c}_l)\right] \tag{2.36}$$

where Γ is a diagonal tensor and $\mathbf{R}$ is a rotation matrix. There are two parameters - the width along the bond direction (rotated so as to be the z-axis), and the width perpendicular to the bond direction. The elements of Γ are defined to be $(1/w_{xy}^2, 1/w_{xy}^2, 1/w_z^2)$.

Finally, additional energy reduction was found for the isolated H_2 molecule by multiplying the orbital by $(1 + \zeta|\mathbf{r} - \mathbf{c}_l|)$, where ζ is a variational parameter.

2.8 Periodic Boundary Conditions

The effects of an infinite system can be approximated by imposing periodic boundary conditions on a finite system. Every particle in the system then has an infinite number of images. Inter-particle distances are calculated using the minimum image convention, which uses only the distance to the closest image.

Care needs to be taken with the wave function when using the minimum image convention. As the inter-particle distance crosses over from one image to another there can be a discontinuity in the derivative of the wave function, leading to a delta function spike in the energy. If this is not accounted for, the VMC energy can become lower than the true ground state because this delta function term in the energy has been neglected. Additionally, the Gaussian orbitals can lower their energy by having a width comparable to or larger than the box size. Then sections of the orbital with large kinetic energy are outside $L/2$, and do not get counted in the integral. This can be fixed by summing over images, or by insuring the wave functions have the correct behavior at $\pm L/2$. We use the latter solution.

The orbitals are multiplied by a cutoff function that ensures its value and first derivative are zero at the box edge. The function we use is

$$f_c(r) = 1 - \exp\left[-\gamma_c (r - r_m)^2\right] \tag{2.37}$$

where r_m is fixed at $L/2$ and γ_c is a variational parameter.

The Jastrow factors are constructed so that they obey the correct cusp conditions as $r \to 0$ and so that the first and second derivatives are zero at $r_m \leq L/2$. The simplest function that satisfies these conditions is a cubic polynomial. Let $y = r/r_m$. Then

$$u(y) = a_1 y + a_2 y^2 + a_3 y^3, \tag{2.38}$$

where $a_1 = (\text{cusp value}) * r_m$, $a_2 = -a_1$, and $a_3 = a_1/3$. Variational freedom is gained by varying r_m, and by adding a general function multiplied by $y^2(y-1)^3$ to preserve the boundary conditions. We choose a sum of Chebyshev polynomials as the general function (**?**). The full Jastrow factor is then

$$u(y) = a_1(y - y^2 + \frac{1}{3}y^3) + y^2(y-1)^3 \sum_i b_i T_i(2y-1), \tag{2.39}$$

where r_m and the b_i are variational parameters. We use five Chebyshev polynomials for the electron-electron part and another five for the electron-nuclear part. We optimized one set of r_m and b_i parameters for all electron-electron pairs in any particular system, and another set of parameters for all electron-nuclear pairs.

Comparisons of the energy and variance of various combinations of forms for the orbital and Jastrow factors are shown in Table 2.5. The variational parameters are given in Tables 2.6 and 2.7. A comparison of the electron-electron Jastrow factors is shown in Figure 2.3. Their short range behavior is similar, but the long range behavior differs between the types of Jastrow factors.

The quality of wave functions is often measured by the percent of correlation energy recovered. For H_2, the HF (no correlation) energy is -1.1336 Hartrees and the exact (full correlation) energy is -1.17447 Ha. **?**) compared a number of forms for electron correlation functions. Their best value recovered 80% of the correlation energy. Using one of these forms, but with better optimization, **?**) recovered 84% of the correlation energy. **?**) obtained 93% of the correlation energy using a Linear Combination of Atomic Orbitals (LCAO) form with 1s, 2s and 2p orbitals, and using the Jastrow factors of **?**).

The variance is higher with those wave functions involving the cubic polynomial, even though the energy is lower. I believe this is mostly likely because the cubic polynomial does not have the correct $1/r$ behavior at large r, but the simple Padé form does. This long range behavior contributes little to the average of the energy, but it contributes more significantly to the variance.

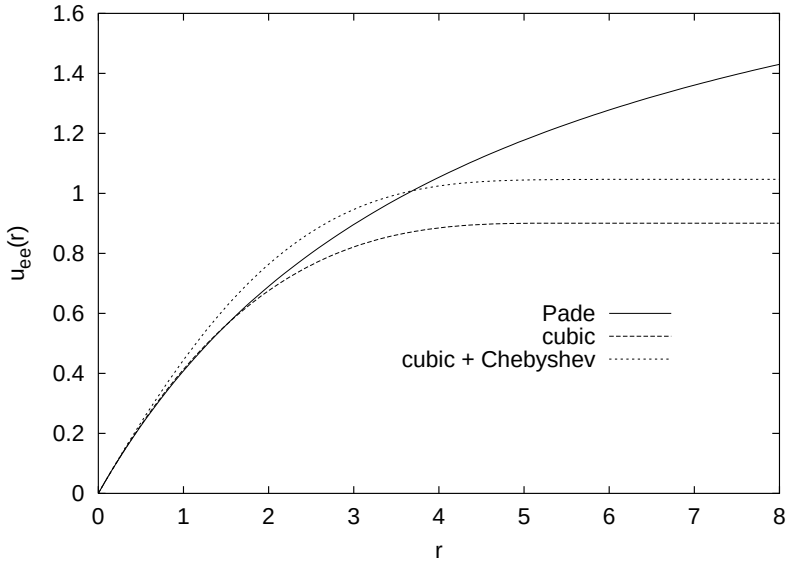


Figure 2.3: Optimized electron-electron Jastrow factor for different forms.

Table 2.5: Comparison of energies and variances for various forms for orbitals and Jastrow factors for a single H_2 molecule.

Ψ	Orbital	Jastrow	Energy	Variance	% CE
A	Isotropic	simple Padé	$-1.1598(4)$	0.046	64.0(9)
B	Anisotropic	simple Padé	$-1.1643(2)$	0.040	75.0(6)
C	Anisotropic + ζ	simple Padé	$-1.1653(2)$	0.033	77.5(5)
D	Anisotropic + ζ	cubic	$-1.1688(2)$	0.039	86.1(6)
E	Anisotropic + ζ	cubic+Chebyshev	$-1.1702(2)$	0.046	89.6(5)

Table 2.6: Values of variational parameters for H_2.

Ψ	Orbital parameters	Jastrow parameters
A	$w = 2.74$	$\beta = 9.913$
B	$w_{xy} = 2.514$, $w_z = 2.977$	$\beta = 10.002$
C	$w_{xy} = 2.416$, $w_z = 2.833$, $\zeta = 0.0445$	$\beta = 9.958$
D	$w_{xy} = 2.357$, $w_z = 2.628$, $\zeta = 0.248$	e-e $r_m = 5.404$
		e-n $r_m = 5.376$

Table 2.7: Values of variational parameters for wave function E

Component	Parameters		
Orbital	$w_{xy} = 2.299$	$w_z = 2.515$	$\zeta = 0.301$
Electron-electron Jastrow	$r_m = 6.281$	$b_0 =$ -1.012	$b_1 = 0.193$
	$b_2 = 0.619$	$b_3 = 0.025$	$b_4 = 0.138$
Electron-nuclear Jastrow	$r_m = 5.329$	$b_0 =$ -2.084	$b_1 = 0.153$
	$b_2 = 0.952$	$b_3 = 1.217$	$b_4 = 1.027$

Chapter 3

Energy Difference Methods

Very often it is the difference in energy between two systems that is of interest, and not the absolute energy of a single system. For a quantity such as the binding energy, we want the difference between the energy of the molecule and the energy of the free atoms. In our CEIMC simulations, we want the change in energy from moving a few nuclei. In VMC optimization, we want to know the change in energy from modifying some of the wave function parameters.

Correlated sampling methods can provide a more efficient approach to computing these energy differences. But the widely used reweighting method has some drawbacks. We will introduce a new method that alleviates some of the drawbacks of reweighting while retaining its advantages.

3.1 Direct Difference

The simplest, and most straightforward way of computing the difference in energy between two systems is to perform independent computations of the energy for each system. Then the energy difference and error estimate are simply

$$\Delta E = E_1 - E_2 \tag{3.1}$$

$$\sigma(\Delta E) = = \sqrt{\sigma_1^2 + \sigma_2^2} \tag{3.2}$$

This method is simple and robust, but has the drawback that the error is related to the error in computing a single system. If the systems are very similar, either in variational parameters or in nuclear positions, the energy difference is likely to be small and difficult to resolve, since σ_1 and σ_2 are determined by the entire system. Similarities between the systems can be exploited with correlated sampling.

3.2 Reweighting

Reweighting is the simplest correlated sampling method.

$$\begin{aligned}\Delta E &= E_1 - E_2 \\ &= \frac{\int dR\ \psi_1^2\ E_{L1}}{\int dR\ \psi_1^2} - \frac{\int dR\ \psi_2^2\ E_{L2}}{\int dR\ \psi_2^2} \\ &= \frac{\int dR\ \psi_1^2\ E_{L1}}{\int dR\ \psi_1^2} - \frac{\int dR\ \psi_1^2 \left(\frac{\psi_2^2}{\psi_1^2}\right)\ E_{L2}}{\int dR\ \psi_1^2\ \left(\frac{\psi_2^2}{\psi_1^2}\right)}\end{aligned} \tag{3.3}$$

An estimate of ΔE for a finite simulation is

$$\Delta E \approx \frac{1}{N} \sum_{R_i \in \psi_1^2} \left[E_{L1}(R_i) - \frac{w(R_i) E_{L2}(R_i)}{\sum_i w(R_i)} \right] \tag{3.4}$$

where $w = \psi_2^2/\psi_1^2$. The same set of sample points is used for evaluating both terms.

Reweighting works well when ψ_1 and ψ_2 are not too different, and thus have large overlap. As the overlap between them decreases, reweighting gets worse due to large fluctuations in the weights. This effect can be quantified by computing the effective number of points appearing in the sum in Eq. (3.4), which is

$$N_{\text{eff}} = \frac{\sum_i w_i^2}{(\sum_i w_i)^2} \tag{3.5}$$

Eventually, one or a few large weights will come to dominate the sum, and the effective number of points will be very small, and the variance in ΔE will be very large.

Particularly pernicious is the case when the nodes differ between the two systems. The denominator of the weight can easily be very small, causing a very large weight value. This is encountered when using reweighting to optimize orbital parameters in VMC (**?**).

In Eq. (3.4) we derived reweighting by drawing points from ψ_1 and computing the properties of both systems from them. It could also be derived by drawing points from ψ_2 as well. We can compute the energy difference both ways and take the average. This gives us the symmetrized reweighting method,

$$\begin{aligned}\Delta E = \quad & \frac{1}{2N} \sum_{R_i \in \psi_1^2} \left[E_{L1}(R_i) - \frac{w_x(R_i) E_{L2}(R_i)}{\frac{1}{N}\sum_i w_x(R_i)} \right] \\ + \quad & \frac{1}{2N} \sum_{R_i \in \psi_2^2} \left[\frac{w_y(R_i) E_{L1}(R_i)}{\frac{1}{N}\sum_i w_y(R_i)} - E_{L2}(R_i) \right]\end{aligned} \tag{3.6}$$

where $w_x = \psi_2^2/\psi_1^2$ and $w_y = \psi_1^2/\psi_2^2$.

3.3 Bennett's Method for Free Energy Differences

First let us digress to discuss computation of the normalization integral. It was mentioned earlier that the Metropolis sampling method makes it difficult to extract information about the normalization integral, which the partition function in the classical case. **?**) demonstrated a method for finding the free energy difference between two systems. We will describe his method in terms of a ratio of normalizations.

We can compute the ratio of two normalizations, Q_1 and Q_2, in a fashion very similar to reweighting.

$$\begin{aligned} Q_1/Q_2 &= \int dR\, \psi_1^2(R) \Big/ \int dR\, \psi_2^2(R) && (3.7)\\ &= \int dR\, \psi_2^2(R) \frac{\psi_1^2(R)}{\psi_2^2(R)} \Big/ \int dR\, \psi_2^2(R) && (3.8)\\ &\approx \sum_{R_i \in \psi_2} \frac{\psi_1^2(R_i)}{\psi_2^2(R_i)} && (3.9) \end{aligned}$$

This is a one-sided estimate, because it only uses samples from system two to compute properties of system one. Note that this sum is the same as the sum of the weights used in reweighting in Eq. (3.4).

Bennett improved on this one-sided estimate, starting with an identity written as

$$Q_1/Q_2 = \frac{Q_1 \int dR\, \psi_2^2\, W \psi_1^2}{Q_2 \int dR\, \psi_1^2\, W \psi_2^2} \tag{3.10}$$

where W is an arbitrary weight function. He found the optimum W by minimizing the variance of the free energy difference,

$$W \propto \frac{1}{Q_1 \psi_2^2 + Q_2 \psi_1^2} \tag{3.11}$$

Let $Q = Q_1/Q_2$. Inserting Eq. (3.11) into Eq. (3.10), we get

$$1 = \frac{\int dR\, \frac{\psi_2^2}{Q_2} \psi_1^2 / \left(\psi_2^2 + \psi_1^2/Q\right)}{\int dR\, \frac{\psi_1^2}{Q_1} \psi_2^2 / \left(Q\psi_2^2 + \psi_1^2\right)} \tag{3.12}$$

Let x represent the configurations sampled from ψ_1 and y the configurations sampled from ψ_2. The finite sample version of this equation is

$$\sum_i \frac{Q\psi_2^2(x_i)}{\psi_1^2(x_i) + Q\psi_2^2(x_i)} = \sum_i \frac{\psi_1^2(y_i)}{\psi_1^2(y_i) + Q\psi_2^2(y_i)} \tag{3.13}$$

The value of Q can be found by a simple iteration

$$Q_{n+1} = Q_n \left[\frac{\sum_y \frac{\psi_1^2(y_i)}{\psi_1^2(y_i)+Q_n\psi_2^2(y_i)}}{\sum_x \frac{Q\psi_2^2(x_i)}{\psi_1^2(x_i)+Q_n\psi_2^2(x_i)}} \right] \tag{3.14}$$

The iteration is started with $Q_0 = 1$ and stopped when the correction factor in brackets is sufficiently close to one. Typically convergence takes less that ten iterations, but if Q is much larger or smaller than one it can take more iterations.

We have written these formulas assuming that the number of sample points from each system is the same. Bennett derived them for case with differing numbers of samples in each sum, and found the best variance was usually very near an equal ratio of computer time spent on each system. In our case the systems are of equal complexity, so this means using equal numbers of points is optimal, or very nearly so.

By properly combining information from both systems, we can get a much better (lower variance) estimate of the ratio of their normalizations than if we had used information from only a single system (one-sided sampling).

3.4 Two-Sided Sampling

We can apply this notion to computing the energy difference between two quantum systems. Consider sampling from some distribution that contains information about both ψ_1 and ψ_2. The simplest such distribution is

$$P = \frac{1}{2}\left[\frac{\psi_1^2}{Q_1} + \frac{\psi_2^2}{Q_2}\right] \tag{3.15}$$

The energy difference can be written as

$$\begin{aligned} \Delta E &= \frac{\int dR\, \psi_1^2 E_{L1}}{Q_1} - \frac{\int dR\, \psi_2^2 E_{L1}}{Q_2} \\ &= \int dR\, P \left(\frac{\psi_1^2 E_{L1}}{Q_1 P}\right) - \int dR\, P \left(\frac{\psi_2^2 E_{L2}}{Q_2 P}\right) \end{aligned} \tag{3.16}$$

In the finite case, we have

$$\Delta E \approx \frac{1}{2N} \sum_{R_i \in x,y} \frac{\psi_1^2(R_i) E_{L1}(R_i)}{Q_1 P} - \frac{\psi_2^2(R_i) E_{L2}(R_i)}{Q_2 P} \tag{3.17}$$

It is important to note that the sum covers samples taken from both ψ_1 and ψ_2. The sum includes both "direct" terms (eg. ψ_1 and E_{L1} evaluated at configurations sampled from ψ_1) and "cross" terms (eg. ψ_1 and E_{L2} evaluated at configurations sampled from ψ_2).

The denominator of the first term in Eq. (3.17) is

$$Q_1 P = \frac{1}{2}\left[\psi_1^2 + \frac{Q_1}{Q_2}\psi_2^2\right] \tag{3.18}$$

The ratio $Q = Q_1/Q_2$ is computed by the Bennett method. The denominator of the second term can be computed similarly.

One major feature of the two-sided method is that it reproduces reweighting in the large overlap regime, and the direct method in the low overlap regime. In the intermediate regime, it smoothly joins the two limits.

To show this, first consider the case where the two systems are very different and the wave functions have low overlap. Here $\psi_1^2(y_i)$ and $\psi_2^2(x_i)$ will be small. Expand Eq. (3.17) into its four terms

$$\begin{aligned}\Delta E = & \underbrace{\frac{1}{2N}\sum_x \frac{\psi_1^2(x_i)E_{L1}(x_i)}{\frac{1}{2}\left[\psi_1^2(x_i)+Q\psi_2^2(x_i)\right]} - \frac{1}{2N}\sum_y \frac{\psi_2^2(y_i)E_{L2}(y_i)}{\frac{1}{2}\left[\psi_1^2(y_i)/Q+\psi_2^2(y_i)\right]}}_{\text{direct}} \\ + & \underbrace{\frac{1}{2N}\sum_y \frac{\psi_1^2(y_i)E_{L1}(y_i)}{\frac{1}{2}\left[\psi_1^2(y_i)+Q\psi_2^2(y_i)\right]} - \frac{1}{2N}\sum_x \frac{\psi_2^2(x_i)E_{L2}(x_i)}{\frac{1}{2}\left[\psi_1^2(x_i)/Q+\psi_2^2(x_i)\right]}}_{\text{cross}}\end{aligned} \tag{3.19}$$

Each denominator will have one large term ($\psi_1^2(x_i)$ or $\psi_2^2(y_i)$) and one small term ($\psi_1^2(y_i)$ or $\psi_2^2(x_i)$). The value of Q is moderate compared to the wave functions, so it will not affect the relative sizes of these terms. Always having one large term in the denominator means there will never be any excessively large contributions to the sum resulting from division by a small value, as happens in reweighting. The cross terms have a small value ($\psi_1^2(y)$ or $\psi_2^2(x)$) in the numerator, and so vanish. The large terms in the denominators in the direct terms cancel the ψ^2's in the numerator, and we are left with

$$\Delta E \approx \frac{1}{N}\sum_x E_{L1}(x_i) - \frac{1}{N}\sum_y E_{L2}(y_i) \tag{3.20}$$

which is just the direct method.

Now for the case where the systems are very similar and have large overlap. Recall from Eq. (3.9) that we can write one-sided estimates for Q as

$$Q = \frac{1}{N}\sum_y w_y(y_i) = 1\Big/\frac{1}{N}\sum_x w_x(x_i) \tag{3.21}$$

where $w_y = \psi_1^2(y_i)/\psi_2^2(y_i)$ and $w_x = \psi_2^2(x_i)/\psi_1^2(x_i)$. Write the four terms of Eq. (3.17) in a different order

$$\begin{aligned}\Delta E = & \frac{1}{N}\sum_x \frac{E_{L1}(x_i)}{1+Qw_x(x_i)} - \frac{Q}{N}\sum_x \frac{w_x(x_i)E_{L2}(x_i)}{1+Qw_x(x_i)} \\ + & \frac{1}{QN}\sum_y \frac{w_y(y_1)E_{L1}(y_i)}{1+w_y(y_i)/Q} - \frac{1}{N}\sum_y \frac{E_{L2}(y_i)}{1+w_y(y_i)/Q}\end{aligned} \tag{3.22}$$

Approximate the denominator of each term by two, replace the leading Q's with the appropriate one-sided approximation, and we get

$$\begin{aligned}\Delta E \approx \quad & \frac{1}{2N}\sum_x \left[E_{L1}(x_i) - \frac{w_x(x_i)E_{L2}(x_i)}{\frac{1}{N}\sum_x w_x(x_i)} \right] \\ + \quad & \frac{1}{2N}\sum_y \left[\frac{w_y(y_i)E_{L1}(y_i)}{\frac{1}{N}\sum_x w_y(y_i)} - E_{L2}(y_i) \right]\end{aligned} \tag{3.23}$$

which is the symmetric version of reweighting given in Eq. (3.6).

Due to computational considerations, it is useful to divide Eqns (3.13) and (3.17) by ψ_1^2 or ψ_2^2 as appropriate, and work with the resulting ratios $w_x = \psi_2^2/\psi_1^2$ and $w_y = \psi_1^2/\psi_2^2$, as was done in Eq (3.22). The values of the wave functions can easily over or under flow double precision variables. It is best to use the log of the wave function, take differences, and then exponentiate. Furthermore, an arbitrary normalization of the wave functions makes no physical difference, but can result in very large or small numbers, even after taking the difference of the logarithms. This problem is ameliorated by subtracting the average value of the log of the wave function from the individual values. Sometimes this is not enough, however, and the value of the energy difference exceeds the range representable in a double precision variable, indicated by NaN (Not a Number). In this case, the overlap is clearly very small and the two-sided method should give the same results as the direct method. The program checks for the energy difference being NaN, and if so, it substitutes the direct method result (the data collected for the two-sided method is a superset of that needed for the direct method). Having done this, the subroutine computing the two-sided energy difference will always return a reasonable answer, an important consideration for a core routine in a program.

3.5 Examples

The first example is of two H_2 molecules in a parallel orientation as shown in Figure 3.1. The bond lengths are at equilibrium, 1.4 Bohr, and the starting separation between the molecules is $d = 2.5$ Bohr.

The energy difference between that configuration and configurations with other intermolecular distances was computed using the direct method, the two-sided method, and reweighting. The resulting energy differences are shown on the left in Figure 3.2. Note that reweighting gets the wrong answer at large separations. This is most likely due to a finite sample size bias. More important is the error in that energy difference, shown on the right in Figure 3.2. Note that both reweighting and the two-sided method have errors that

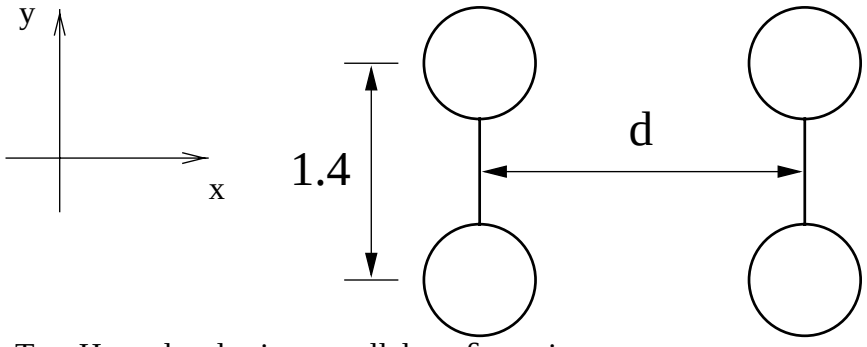


Figure 3.1: Two H_2 molecules in a parallel configuration

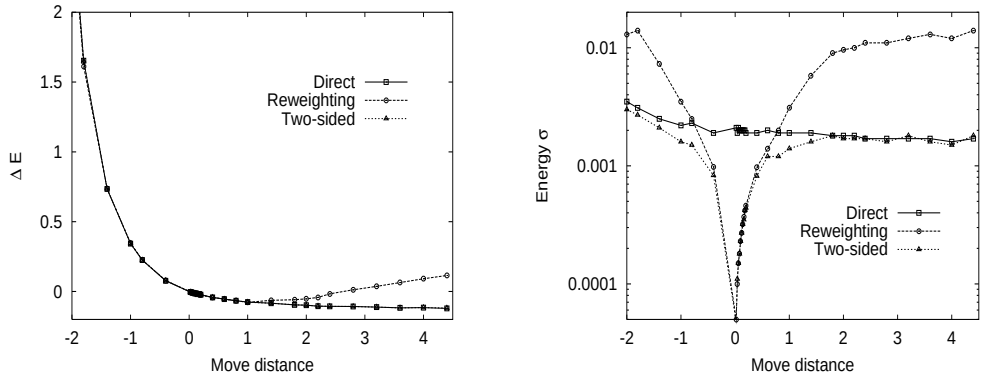


Figure 3.2: Energy difference (left) and the estimated statistical error (on logscale) (right) for two H_2 molecules in a parallel configuration, starting from d=2.5 Bohr.

drop to zero as the overlap increases. This graph also clearly shows the properties of the two-sided method mentioned previously, behaving like reweighting at small changes in the separation (large wave function overlap), and smoothly crossing over to the direct method for at large changes in the separation (small wave function overlap).

Reweighting and the two-sided method may give biased results because there are a finite number of sample points in the sums in Eqns. (3.6) and (3.17). To test for this, a sum of a given length is repeated many times and the average energy difference for that length computed. The test for a bias was performed on a Li_2 molecule. The energy difference was computed between a bond length of 4.5 Bohr and the equilibrium bond length of 5.05 Bohr. Figure 3.3 shows the results for different numbers of points in the sum. Reweighting shows a much larger finite sample size bias than the two-sided method, which has almost none.

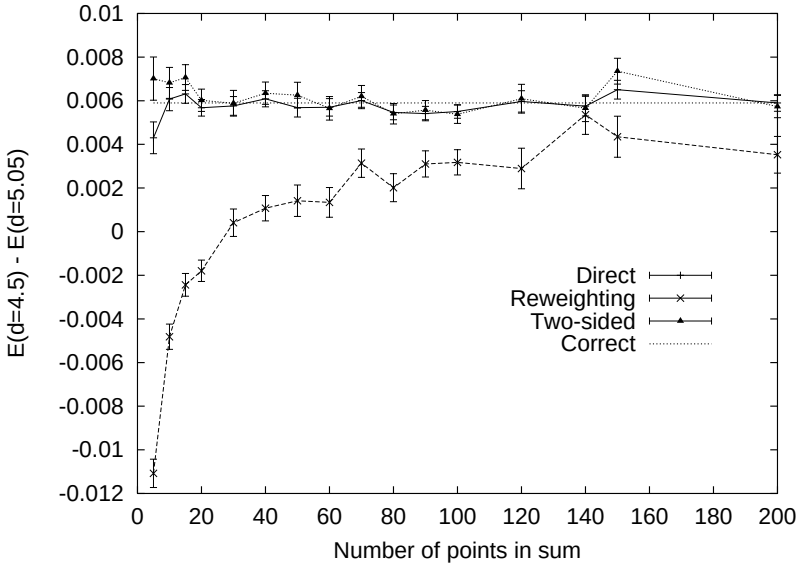


Figure 3.3: Finite sample size bias in the energy difference of Li_2.

3.5.1 Diffusion Monte Carlo

Using the two-sided method (or reweighting, for that matter) with DMC is slightly more complicated. The reweighting transformation applied to the basic DMC iteration gives

$$f_1(R';t+\tau) = \int dR\ f(R;t)G_1(R\rightarrow R';\tau) \tag{3.24}$$

$$= \int dR\ f(R';t)G_2(R\rightarrow R';\tau)\frac{G_1(R\rightarrow R';\tau)}{G_2(R\rightarrow R';\tau)} \tag{3.25}$$

The weight $w = G_1/G_2$ must be computed over several iterations. The final weight used in the correlated sampling formulas is a product of the weights of every iteration.

The weight factor is not quite right, due to the rejection step. Since the rejection ratio for DMC is very small ($< 1\%$), ignoring the issue should not introduce a large error. **?**) give a more sophisticated method for dealing with rejections.

The fixed-node condition also has to be obeyed, and configurations that cross a node while projecting have their weight set to zero.

A version of reweighting was implemented by **?**) as the differential Green's function Monte Carlo method (actually DMC). He used the response to an external field to determine the dipole moment of LiH. The same trial function was used, so the drift term was the same between both systems. Only the branching term was different; that entered as a weight factor. In our case, the trial function and the nuclear positions may be different between the two systems.

The top of Figure 3.4 shows the difference in DMC energies using the various methods.

The energy difference was computed starting from the equilibrium bond length of 5.05 Bohr. Partly because of the need to project for several DMC steps, the two-sided method has a fairly small range where it does better than the direct method (compared with the range for VMC). For comparison, the VMC results are shown at the bottom of Figure 3.4. There are two lines in the DMC graph for the direct method. The implementation had a limitation where only one projection to accumulate the weights would occur at a time. We used 30 steps in the projection, and consequently could only get the weights once every 30 steps. The limited data line is computed from data collected once every 30 steps (the same amount of data available to the correlated methods) and is the line the two-side method joins on to. The full data line used all the data available in the simulation and so has a lower statistical error.

3.5.2 Binding Energy

To compute the binding energy, let the non-interacting system be ψ_2, and the fully interacting system be ψ_1. The nuclear positions are the same for both systems, and the appropriate interaction terms are set to zero for the non-interacting system.

A pair of H_2 molecules in a parallel configuration was used as the test system. The binding energy we are interested in is that of the interacting molecules minus separate H_2 molecules (and not the separate atoms).

$$E_B = E\left((H_2)_2\right) - 2E(H_2) \tag{3.26}$$

There is a problem in that the electrons in the fully interacting system can switch molecules and have no effect on the computation, but these configurations are very unlikely in the non-interacting system. This leads to an artificially small overlap. The solution in this symmetric case is to restrict the domain of integration. The electron coordinates are ordered along the x-axis so that $x_1 < x_2$ and $x_3 < x_4$.

To see that this restriction is exact, consider the integral

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \; f(|x_1 - x_2|) g(x_1, x_2) \tag{3.27}$$

where f corresponds to the electron-electron Jastrow factor and g is symmetric under the interchange of x_1 and x_2 and corresponds to the electron-nucleus Jastrow factor and the square of the Slater determinant. Change variables to $R = (x_1 + x_2)/2$ and $r = x_1 - x_2$. Now we have

$$\int_{-\infty}^{\infty} dR \int_{-\infty}^{\infty} dr \; f(|r|) g(R + r/2, R - r/2) \tag{3.28}$$

The integral over r is even, and we only need to integrate over half of the interval ($r < 0$ or $r > 0$), which corresponds to the restrictions $x_1 < x_2$ or $x_1 > x_2$.

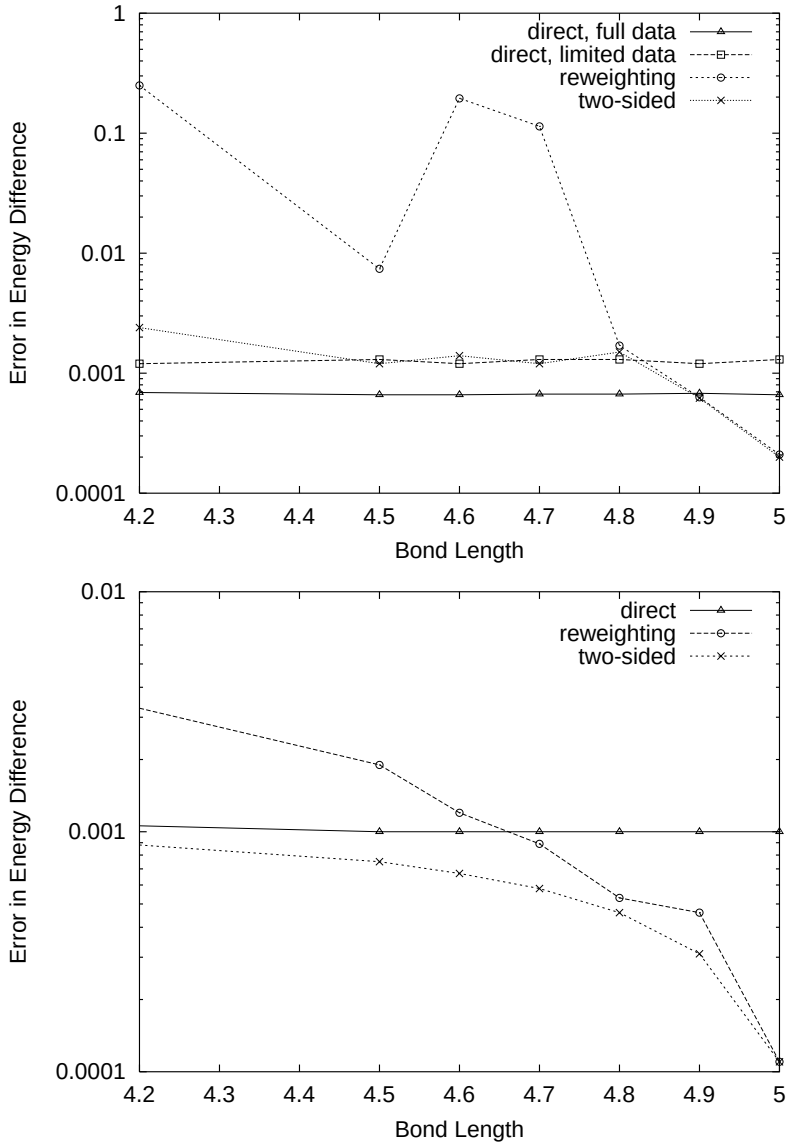


Figure 3.4: Error in energy difference of Li_2 using DMC (top) and VMC (bottom)

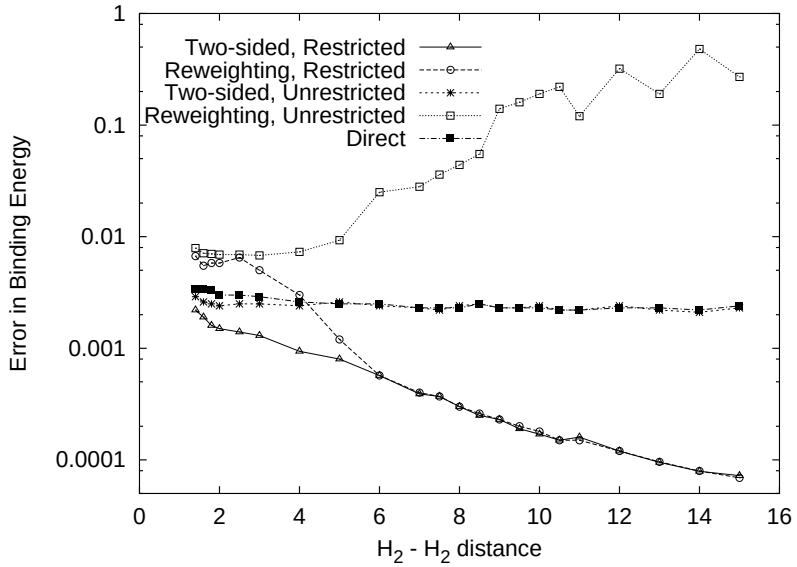


Figure 3.5: Error in VMC binding energy of H_2-H_2 system

Figure 3.5 shows the error in the VMC binding energy for various intermolecular distances. Without restricting the domain of integration, reweighting performs poorly, and the two-sided method reproduces the results of the direct method. With the restricted domain, the correlated methods perform quite well.

Chapter 4

Wave Function Optimization

Variational Monte Carlo (VMC) depends crucially on the optimization of parameters in the wave function to find the minimum energy. The general problem of function optimization is a well-studied area. For a general introduction to various optimization techniques, see **?**). For more in-depth work, consult **?**), **?**), or **?**).

The main difficultly in applying these techniques to optimizing VMC wave functions is noise - we only get stochastic estimates for function values or gradients. **?**) describes several strategies for optimization in the presence of noise. We will divide these into three categories.

The first strategy is to convert the problem into a nearby smooth, non-noisy problem, and solve that problem instead. Fixed sample reweighting takes this approach by sampling some set of configurations and optimizing with just these configurations.

The second approach is to reduce the noise to negligible levels, and proceed with regular optimization techniques. This is possible with a Newton method, where the first and second derivatives of the function are computed, and the number of iterations needed for convergence hopefully is small.

The third approach is to use a method tailored to handle noise. The Stochastic Gradient Approximation (SGA) is such a method. Also somewhat in this category, we will examine a method that is essentially a biased random walk, and the moves are accepted or rejected based on the whether or not the energy decreases.

These approaches will be compared on several problems of different sizes to see how they scale. We will use a single H_2 molecule, and collections of 8,16, and 32 H_2 molecules in a box as trial problems.

4.1 Energy vs. Variance Minimization

There is a choice of objective functions - either the energy or the variance of the energy can be minimized. Under certain circumstances, variance minimization is more stable that energy minimization. For the reweighting method, this is definitely true, but it may not be the case for the other optimization techniques. It is generally held that variance optimization would produce better values for observables other than energy (**?**), but this may not always be the case (**??**). The argument is that the variance is more sensitive to parts of the wave function that do not contribute to the energy. As we have seen in Chapter 2, having incorrect long range behavior in the H_2 wave function does not affect the energy much, but does cause the variance to rise. In other words, variance minimization should yield a "smoother" wave function, which should then have better non-energy properties.

4.2 Fixed Sample Reweighting

Fixed sample reweighting with minimization of the variance was popularized by **?**), and has been used extensively since then. The current state of the art is described by Kent (**??**).

The core of the method is the single sided reweighting method described in Chapter 3. A number of configurations are sampled from a distribution with variational parameters α_0. The energy at an arbitrary value of the variational parameter, α, is computed by

$$E(\alpha) = \sum_i w(R_i;\alpha) E_L(R_i;\alpha) / \sum_i w(R_i;\alpha) \tag{4.1}$$

where $w(R_i;\alpha) = \psi^2(R_i;\alpha)/\psi^2(R_i;\alpha_0)$. Alternatively, one could compute the variance by

$$A(\alpha) = \sum_i w(R_i;\alpha) (E_L(R_i;\alpha) - E_T)^2 / \sum_i w(R_i;\alpha) \tag{4.2}$$

where E_T can either be the weighted average energy (4.1) or it could be a guess at the desired energy.

The weights in these expressions can get very large when the variational parameters move far from the sampled value α_0, and especially when the parameters that affect the nodes are adjusted. Then just a few configurations will dominate the sum, and the energy estimator can often give meaningless low values. The variance estimator, however, will remain more stable in this situation. For either estimator, the best fix is to regenerate the configurations being used when the parameters move too far away from α_0. This can be used in conjunction with the enhancements described below.

A second advantage of the variance estimator is that the weights can be modified without changing the location of the minimum (**?**). The same is not true for the energy. The

problem of a few large weights can be solved by limiting them to a maximum value (**?**), or more simply by just setting them all to one (**?**). **?**) tame the fluctuating weights by sampling from a positive definite guiding function. In this cases, E_T should be set to a best guess, or slightly below, because the energy estimator will not be reliable.

Further increases in stability can be gained by limiting outliers in Eq. (4.2). Large outlying values have a disproportionately large effect on the variance, but their contribution is not that meaningful. **?**) gives a procedure choosing a cutoff that will reduce its effect as the number of samples increases. We used a simpler approach, removing from the sum any values greater than 5 standard deviations from the average.

Another efficiency improvement can be exploited when the Jastrow U is linear in the variational parameters. Then the variational parameters can be factored out of the sum over interparticle distances, and the value of that sum can be stored. Fixed sample reweighting has been applied mostly to optimizing Jastrow factors, and not parameters in the Slater determinant, so this results in a dramatic time savings. In our case we have both Jastrow and determinantal parameters, and the code spends about 40% of its time computing the Jastrow factor. This percentage will decrease as the system size increases, since the Jastrow computation is $O(N^2)$ but the determinantal part requires matrix work that is $O(N^3)$. We did not implement this improvement, so bear in mind when perusing the results that the reweighting time could be reduced, probably by 30%.

An additional advantage of reweighting is that, since it is solving a smooth problem, an off-the-shelf minimizer can be used. We used the DSMNF general minimizer from the PORT library, which uses only the function values and does not need any derivatives. Routines to minimize sums of squares are also available, but we did not try them. The fixed sample reweighting algorithm is then: Generate a set of configurations and minimize the variance with this set. Generate a new set of configurations using the new variational parameters and find the minimum variance again. Repeat for several steps to ensure convergence.

4.3 Newton Method

The Newton method makes use of the first and second derivatives. We can approximate a function near its minimum as a quadratic surface

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)^T \cdot \mathbf{b} + (\mathbf{x}-\mathbf{x}_0)^T \cdot \mathbf{A} \cdot (\mathbf{x}-\mathbf{x}_0) \tag{4.3}$$

where $b_i = \frac{\partial f}{\partial x}$ and $A_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ is the Hessian matrix. The location of the minimum is then given by

$$\mathbf{x}_0 = \mathbf{x} - \mathbf{A}^{-1} \cdot \mathbf{b} \tag{4.4}$$

Since we are likely to start in a region where f is not quadratic, this step is iterated several times.

This procedure, along with analytic evaluation of the derivatives, was applied to VMC energy minimization by **?**). Analytic derivatives of the local energy with respect to determinantal parameters are given by **?**), but these were used in the context of a reweighting minimization.

Recall the VMC energy is computed by

$$E = \langle E_L \rangle = \int dR\ \psi^2(\alpha) E_L(\alpha) \Big/ \int dR\ \psi^2(\alpha) \tag{4.5}$$

We want the derivatives of E with respect to various variational parameters, α. These could be computed with finite differences and reweighting, but it is better to do some analytical work on this expression first.

?) use some Green's relations to eliminate explicit derivatives of the local energy (**?**), and derive the following expression for the gradient,

$$\frac{\partial E}{\partial \alpha_m} = 2\left[\langle E_L\ \psi'_{\ln,m}\rangle - \langle E_L\rangle\langle \psi'_{\ln,m}\rangle\right] \tag{4.6}$$

where

$$\psi'_{\ln,m} = \frac{\partial \ln\psi}{\partial \alpha_m} = \frac{1}{\psi}\frac{\partial \psi}{\partial \alpha_m}. \tag{4.7}$$

They also give the expression for the Hessian,

$$\begin{aligned}\frac{\partial^2 E}{\partial \alpha_m \partial \alpha_n} &= 2\Big\{\langle E_L\ \psi''_{\ln,m,n}\rangle - \langle E_L\rangle\langle \psi''_{\ln,m,n}\rangle \\ &\quad +2\left[\langle E_L\ \psi'_{\ln,m}\psi'_{\ln,n}\rangle - \langle E_L\rangle\langle \psi'_{\ln,m}\psi'_{\ln,n}\rangle\right] \\ &\quad -\langle \psi'_{\ln,m}\rangle\frac{\partial E}{\partial \alpha_n} - \langle \psi'_{\ln,n}\rangle\frac{\partial E}{\partial \alpha_m} \\ &\quad +\langle \psi'_{\ln,m}E'_{L,n}\rangle\Big\}\end{aligned} \tag{4.8}$$

where

$$E'_{L,n} = \frac{\partial E_L}{\partial \alpha_n} \tag{4.9}$$

and

$$\psi''_{\ln,m,n} = \frac{\partial^2 \ln\psi}{\partial \alpha_m \partial \alpha_n}. \tag{4.10}$$

Computing the first derivatives of the wave function analytically is relatively easy. Computing the second derivatives with respect to parameters in the Jastrow factor is also easy analytically. For parameters that appear in the determinant, however, second derivatives are more difficult. For this reason we compute most of the first derivatives analytically, and use these to compute the second derivatives with a simple finite difference scheme. The first derivative of the local energy was computed with finite differences. The derivative of the orbital cutoff parameter, which is the same for all the orbitals, was also computed with finite differences.

An advantage of the Newton approach over the gradient-only approaches is that it has information about how big of step should be taken, whereas the step size is a parameter that must be tuned in the gradient-only methods. The drawback, though, is a greater sensitivity to noise. The gradient and Hessian must be sufficiently accurate, or the Newton iteration will get wildly wrong results. More precisely, it is the non-linear process of taking the inverse in Eq. (4.4) that causes the problem. Furthermore, this sensitivity to noise increases with the number of parameters.

Another problem is parameter degeneracy, or near degeneracy. This will make the Hessian singular, or nearly so. Even if it not exactly singular, being nearly singular is the equivalent of dividing by a small number, which will also greatly magnify the effects of noise. The usual solution is use of the Singular Value Decomposition (SVD). See **?**) or **?**) for a description of the algorithm. A more detailed look at "regularization" (of which the SVD is one method) is in **?**). The SVD starts by decomposing a matrix as

$$\mathbf{A} = \mathbf{PDQ} \tag{4.11}$$

where **P** and **Q** are unitary matrices and **D** is a diagonal matrix. The elements of **D** are the eigenvalues of $\mathbf{A}^T\mathbf{A}$. For our square, symmetric matrix, these are are the squares of the eigenvalues of **A**. We can also take **P** and **Q** to be the eigenvectors of **A**. The utility of the SVD is seen when we write the inverse of **A** as

$$\mathbf{A}^{-1} = \mathbf{Q}^T\mathbf{D}^{-1}\mathbf{P}^T \tag{4.12}$$

If **A** is singular, then at least one of its eigenvalues is zero. In this case, zero eigenvalues also indicate parameter degeneracy, so it's not really necessary to move in the directions corresponding to the zero eigenvalues. To avoid moving in these these directions, and to stabilize the inverse, set $1/d_i$ in Eq. (4.12) to zero when d_i is smaller than some cutoff.

With the eigenvalue decomposition we have an additional technique - negative eigenvalues correspond to uphill directions and mean we are at a saddle point or are far from a

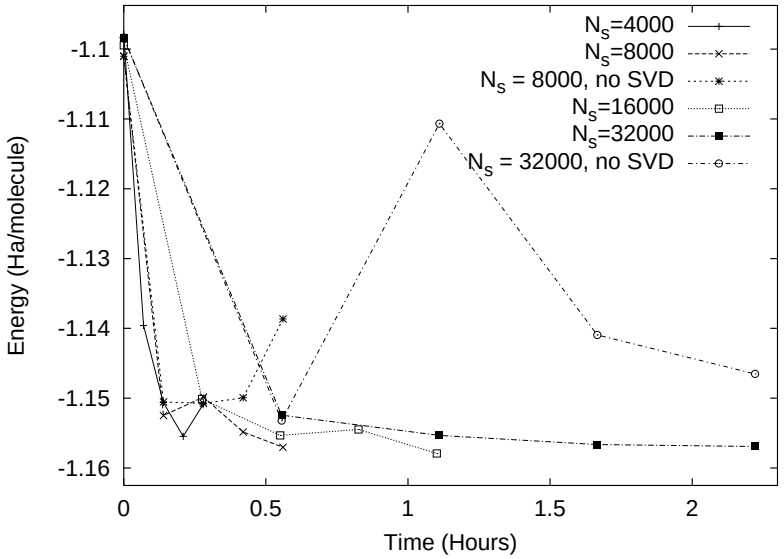


Figure 4.1: Examples using the Newton iteration with varying amounts of noise.

region where the quadratic approximation is good. [1] The simplest way of handling this is to ignore negative eigenvalues. So we remove small positive and all negative eigenvalues when solving Eq. (4.4).

Some examples of this Newton iteration with 8 H_2 molecules are shown in Figure 4.1. N_s is the number of samples used in computing the gradient and Hessian. Unless otherwise noted, the SVD method for solving Eq. (4.4) was used with removal of eigenvalues less than 0.01. Other runs without using regularization are not shown because they diverge very drastically.

4.4 Stochastic Gradient Approximation

The Stochastic Gradient Approximation (SGA) was designed by **?**) to handle optimization with noisy gradients. It was first applied to VMC optimization by **?**).

The SGA iteration can be written as

$$\alpha_i = \alpha_{i-1} - h\gamma_i \nabla_\alpha E(\alpha_{i-1}). \tag{4.13}$$

where h is a step size parameter and γ_i is some specially chosen series.

[1] Much of the complexity of current Newton and quasi-Newton optimization methods is in deciding how to move the parameters when the quadratic approximation is not good. Typically it involves a line minimization in the gradient direction or some sort of back tracking.

There are some conditions on γ_i that must be satisfied in order for this iteration to converge. They are

$$\gamma_i > 0 \tag{4.14}$$

$$\sum_{i=1}^{\infty} \gamma_i = \infty \tag{4.15}$$

$$\sum_{i=1}^{\infty} \gamma_i^2 < \infty \tag{4.16}$$

The condition given by Eq. (4.15) allows the iteration to reach anywhere in parameter space. The condition in Eq. (4.16) is needed so the effects of noise will eventually be damped out. An obvious choice for γ_i is $1/i$. For more discussion on these conditions and for some conditions on the objective function, see **?**) and **?**).

We can analyze the convergence in the limiting case of no noise in one dimension. First let us make a continuous version of Eq. (4.13) by letting $\gamma(t) = dt/t$ and $d\alpha = \alpha(t) - \alpha(t - dt)$. Then in the $dt \to 0$ limit, the SGA iteration is

$$\frac{d\alpha}{dt} = -\frac{h}{t}\nabla_\alpha f(\alpha(t)) \tag{4.17}$$

Now let us assume that f has a quadratic form, $f(\alpha) = \frac{1}{2}A\alpha^2 + B\alpha + f_0$, with a minimum at $\alpha = -B/A$. Now Eq. (4.17) is

$$\frac{d\alpha}{dt} = -\frac{h}{t}\left[A\alpha + B\right] \tag{4.18}$$

The solution is

$$\alpha(t) = -B/A + a_0 t^{-hA} \tag{4.19}$$

where a_0 is a constant of integration. So we see that it will converge to the solution at $t \to \infty$, with a rate that is controlled by the curvature of the potential and our choice of h.

Now consider generalizing to the case where $\gamma(t) = dt/t^\delta$. Our continuous equation is then

$$\frac{d\alpha}{dt} = -\frac{h}{t^\delta}\left[A\alpha + B\right] \tag{4.20}$$

The solution is

$$\alpha(t) = -B/A + a_0 \exp\left[-hAt^{1-\delta}/(1-\delta)\right] \tag{4.21}$$

We see that the smaller δ is, the faster the convergence. If there were no noise, $\delta = 0$ would indeed be the best choice.

Now let us represent the noise in the gradient with an additive noise term, $\eta(t)$. Then Eq. (4.20) is

$$\frac{d\alpha}{dt} = -\frac{h}{t^\delta}\left[A\alpha + B + \eta(t)\right] \tag{4.22}$$

Previously we considered case where noise was negligible. Now consider the case where the noise dominates, so Eq. (4.22) becomes

$$\frac{d\alpha}{dt} = -\frac{\eta(t)}{t^{\delta}} \tag{4.23}$$

The solution is the integral

$$\alpha(t) = -\int^{T} dt \frac{\eta(t)}{t^{\delta}} \tag{4.24}$$

To look at convergence, we need to compute the variance of α integrated over the noise. Take the noise to have a probability distribution $P(x,t)$ with zero mean and variance σ. The variance of α is then

$$\sigma^2(\alpha) = -\int_{-\infty}^{\infty} dx \int^{T} dt\ P(x,t) \frac{x^2}{t^{2\delta}} \tag{4.25}$$

If we take $P(x,t)$ to have no dependence on t, the integrals factor and we get

$$\sigma^2(\alpha) = -\frac{\sigma^2}{1-2\delta}\ \frac{1}{T^{2\delta-1}} \tag{4.26}$$

Here we see that larger values of δ lead to faster convergence of the noise. Since smaller values of δ lead to faster convergence of the non-noisy problem, we need an intermediate value of δ to balance these effects.

One variation, suggested by **?**), is to use $\delta = 1/2$ and use the cumulative average of the variational parameters. This value of δ violates the condition in Eq. (4.16), but this condition is there to insure the noisy part converges. Instead we use the cumulative averaging process to remove the noise.

Another acceleration technique involves monitoring the sign of the gradient (**?**). Far from the minimum the gradient will not often change sign between successive steps. Close to the minimum, the noise will eventually dominate, and the gradient will change sign more often. The acceleration procedure is to only update γ_i when the sign of the gradient changes. This also has the advantage of adjusting the convergence of each parameter separately.

In practice, starting the series at $\gamma_1 = 1$ tends to make the first steps have a dramatically larger effect on the parameters than subsequent steps. Often, the first few steps would move the parameters very far from the minimum, and then the iteration will take a long time to converge. In this work we started the series at $i = 10$ to minimize this effect.

We tried several of these SGA variants on the box of 8 H_2 molecules. We used $h = 3$ when $\gamma_i = 1/\sqrt{i}$ and $h = 10$ when $\gamma_i = 1/i$. This way the initial step sizes (give by $h\gamma_i$) were similar. Figure 4.2 shows the convergence of one of the variational parameters (r_m for the electron-electron Jastrow). The convergence of the energy is also shown. We see that the two accelerated methods converge faster than the simple SGA.

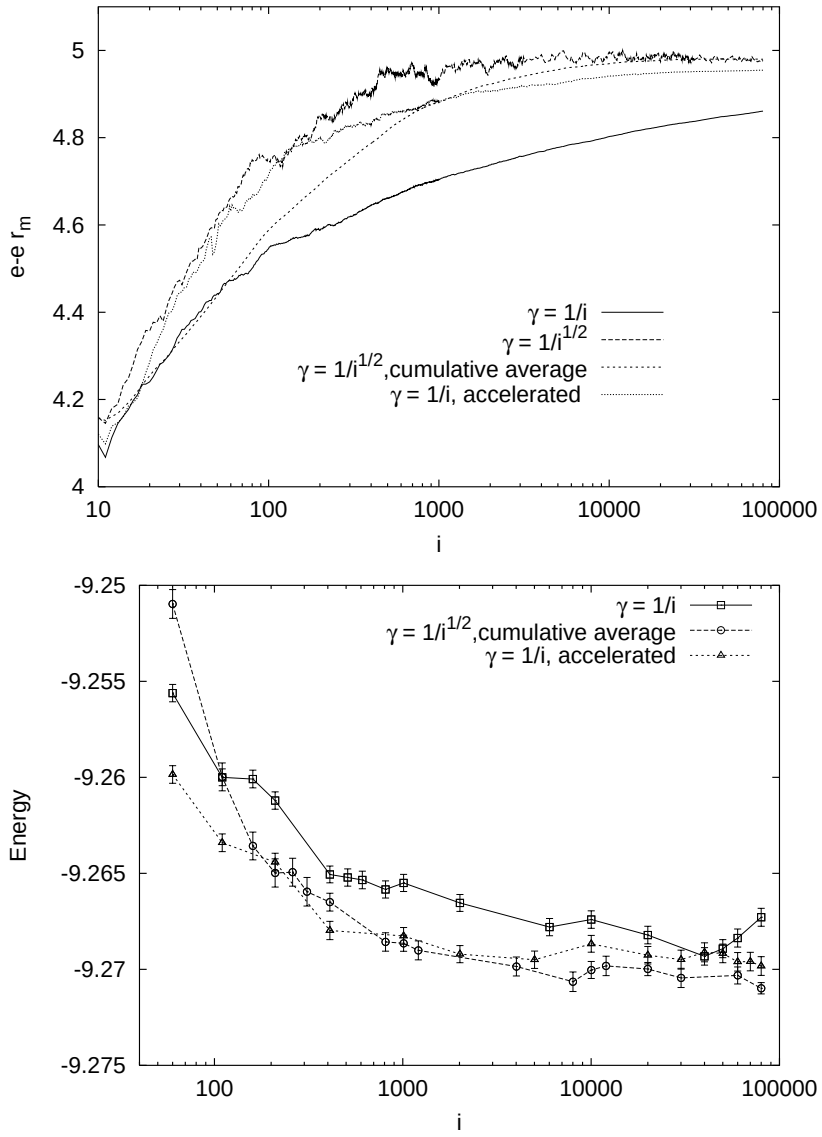


Figure 4.2: Examples of SGA. The graph on the top shows the convergence of one variational parameter for several SGA algorithms. The graph on the bottom shows the resulting energy.

4.5 Gradient Biased Random Walk

We introduce a new method that is made possible by the two-sided energy difference method in Chapter 3. Using this, it is relatively easy to determine whether a change of the trial parameters lowers the energy or not. This determination can be fitted onto a number of methods, even a random walk. We evaluate the gradient and make a trial move in the gradient direction, similar to the SGA. Unlike the SGA, the move is accepted only if it lowers the energy. Since the gradient is noisy, we are effectively making a random walk that is biased in the gradient direction, hence the name Gradient Biased Random Walk (GBRW).

The trial move is

$$\alpha_T = \alpha_{i-1} - h\nabla_\alpha E(\alpha_{i-1}). \tag{4.27}$$

where h is randomly chosen from $[0, h_{\max}]$. To provide some simple adaptivity, $h_{\max}$ is adjusted during the run. If a trial move is rejected, $h_{\max}$ is decreased via multiplication by some factor, usually 0.5 or 0.6. If a trial move is accepted, it is increased by multiplying by the reciprocal of that same factor.

Currently the level of convergence of this method is controlled by how well the energy difference is computed. In other words, once the energy differences are of the same size as the estimated error, it simply fluctuates. There are several possibilities for making a convergent method. The first is to take the cumulative average of the parameters, or add a damping parameter as in the SGA. The second is to increase the number of samples to compute the energy difference (and so decrease the noise) at each iteration.

4.6 Comparison of methods

We test the various optimization methods and compare their run times. The test systems are an isolated H_2 molecule, and 8,16, and 32 molecules in a box at $r_s = 3.0$, a fairly low density. Each system has 12 parameters in the Jastrow factor, and 3 determinantal parameters per molecule, plus one more for the box cutoff (which is the same for all the orbitals). Thus we have 15, 37, 61, and 109 variational parameters, respectively. For the starting parameters,we set the Jastrow cutoff to $r_m = 4.0$, the orbital widths to 2.0, the orbital box cutoff to 1.0, and all the other parameters to zero.

The Newton method used the regularization method with a cutoff of 0.01 for $N = 8, 16$ and a cutoff of 0.1 for $N = 32$. No regularization was used for $N = 1$. The SGA method used $\gamma_i = 1/\sqrt{i}$ and parameter averaging. Reweighting used 16000 configurations for $N = 1$ and 1000 configurations for $N = 8$ and 16. We did not attempt reweighting on the largest system.

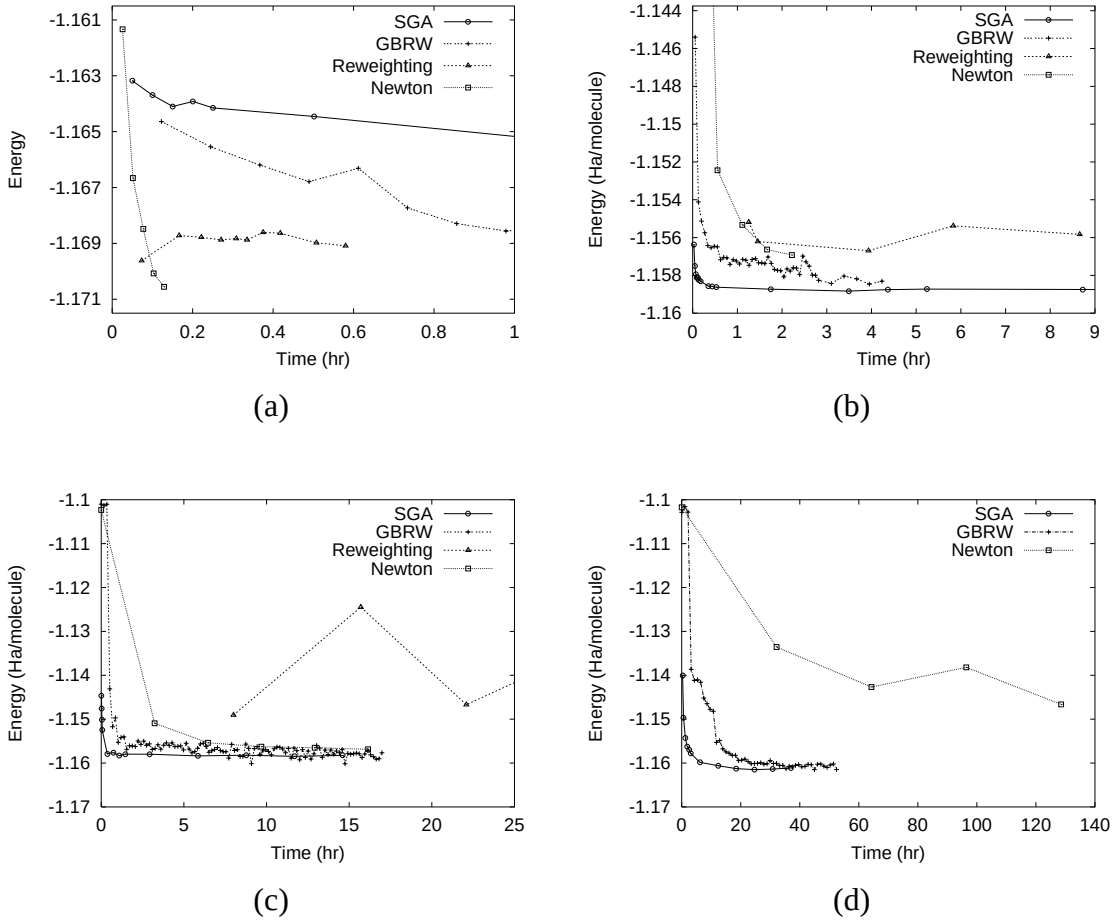


Figure 4.3: Optimization methods applied to (a) Single H_2 (b) 8 H_2's (c) 16 H_2's (d) 32 H_2's

The best way to compare these methods would be to run them all many times starting from different random number seeds. The average of the resulting distribution would give the average quality of each method, and the spread of the distribution would indicate the stability. However, this is time-consuming and instead, as a first approximation we present the results for a single run of each method in Figure 4.3. The times are in hours on an AMD Duron 600 Mhz (which is approximately 1/2 to 2/3 the speed of a 195 Mhz R10000 in an SGI Origin).

For the single molecule, it is clear that the Newton method is the best method. The reweighting method also performs well, and the two gradient methods take longer to converge. As the system size increases, however, the gradient methods do better, with the SGA method doing the best.

The Newton method in particular has difficulty with stability as the system size increases. It needs to be run long enough so the noise is small enough that it does not affect

the results.

The reweighting method performs surprisingly poorly on the larger systems. Looking more closely at the results of reweighting for the $N = 8$ case, we get a total energy of $-9.244(2)$ Ha and a variance of 0.30. From the SGA we get $E = -9.275(3)$ Ha and a variance of 0.42. From the GBRW we get $E = -9.268(2)$ Ha and a variance of 0.36. It appears that the problem is with the variance minimization and we have a case where the minimum variance solution is not the lowest energy solution. Although on the scale of the total energy, the difference between reweighting and the SGA is only 0.3%. On the scale of the correlation energy in the isolated molecule, this difference is about 10%.

4.7 Future Work

We have compared a few basic methods for VMC parameter optimization. Many more improvements and modifications could be conceived and tried.

Currently we ran these with set numbers of iterations and numbers of samples, then looked at the results, and perhaps made adjustments and tried again. What would be very helpful is some sort of adaptivity - adjusting the number of samples or even the type of method as the optimization proceeds in order to ensure convergence.

So far the gradient-only methods seem to have the advantage, but have the disadvantage that they require a step size be set manually. In order to generate a trial step size automatically, a secant updating method could be tried, where successive gradient evaluations are used to build up an approximate inverse Hessian (**?**). These methods are often superior to using the actual Hessian (**?**) , but it is not clear how the presence of noise will affect the algorithm.

Finally, it would be instructive to perform these comparisons on systems containing atoms with higher atomic number.

Chapter 5

Coupled Simulation Methods

There are several issues we have to deal with when constructing an efficient CEIMC simulation. The first is noise from the QMC evaluation of the energy. We will discuss a modification to the Metropolis acceptance ratio, called the penalty method, that will accommodate noise. Next we will examine some of the details involved in a CEIMC simulation, and finally give results for a single H_2 molecule.

5.1 Penalty Method

The Metropolis acceptance ratio, from Chapter 2, is $\min[1, \exp(-\Delta)]$, where $\Delta = \beta[V(s') - V(s)]$. The QMC simulation will yield a noisy estimate for Δ, which we denote as δ

The exponential in the acceptance ratio is nonlinear, so that $\langle \exp(-\delta) \rangle \neq \exp(\langle -\delta \rangle)$. The noise will introduce a bias into our acceptance ratio formula. To avoid this bias in our simulations, we can either run until the noise is negligible, or we can try find a method that tolerates noise.

Typical energy differences for moves in our simulations are on the order of $.01 - .05$ Ha. If we want an error level of 10% (statistical error of .001 Ha) [1] it would take about 7 hours of computer time for a system of 16 H_2 molecules. We need to perform hundreds of these steps as part of the classical simulation, so clearly a method that could tolerate higher noise levels would be very beneficial. The penalty method of **?)** does this, and our simulations run with noise levels on the order of .01 Ha, which only takes about 4 minutes of computer time.

In the penalty method, we start with detailed balance, written as

$$A(s \to s') = A(s' \to s) \exp[-\Delta]. \tag{5.1}$$

[1] The usual error level considered chemical accuracy is 1 kcal/mol = .0016 Ha

To deal with noise, we would like to satisfy detailed balance on average, We introduce an instantaneous acceptance probability, $a(\delta)$, that is a function of the estimated energy difference. The average acceptance probability is the instantaneous one averaged over the noise,

$$A(s \to s') = \int_{-\infty}^{\infty} d\delta P(\delta; s \to s') a(\delta) \tag{5.2}$$

The detailed balance equation we would like to satisfy is then

$$\int_{-\infty}^{\infty} d\delta P(\delta; s \to s') \left[a(\delta) - e^{-\Delta} a(-\delta) \right] = 0 \tag{5.3}$$

Suppose the noise is normally distributed with variance, σ. Then

$$P(\delta) = (2\sigma^2\pi)^{-1/2} \exp\left[-\frac{(\delta-\Delta)^2}{2\sigma^2} \right] \tag{5.4}$$

A simple solution to Eq. (5.3) is

$$a(\delta) = \min\left[1, \exp(-\delta - \frac{\sigma^2}{2}) \right] \tag{5.5}$$

The extra $-\sigma^2/2$ term causes addition rejections of trial moves due to noise. For this reason it is called the penalty method.

To verify that the solution in Eq. (5.5) satisfies detailed balance (5.3), let us compute the average acceptance probability

$$\begin{aligned}
A(\Delta) &= \frac{1}{\sqrt{2\sigma^2\pi}} \int_{-\infty}^{\infty} d\delta\, e^{-(\delta-\Delta)^2/2\sigma^2} \min\left[1, e^{-\delta-\sigma^2/2} \right] \\
&= \frac{1}{\sqrt{2\sigma^2\pi}} \int_{-\infty}^{-\sigma^2/2} d\delta\, e^{-(\delta-\Delta)^2/2\sigma^2} + \frac{1}{\sqrt{2\sigma^2\pi}} \int_{-\sigma^2/2}^{\infty} d\delta\, e^{-(\delta-\Delta)^2/2\sigma^2} e^{-\delta-\sigma^2/2} \\
&= \frac{1}{\sqrt{2\sigma^2\pi}} \int_{-\infty}^{-\sigma^2/2-\Delta} d\delta'\, e^{-\delta'^2/2\sigma^2} + \frac{1}{\sqrt{2\sigma^2\pi}} \int_{\sigma^2/2-\Delta}^{\infty} d\delta''\, e^{-\Delta} e^{-\delta''^2/2\sigma^2} \\
&= \frac{1}{2}\mathrm{erfc}((\sigma^2/2+\Delta)/2\sigma^2) + \frac{1}{2} e^{-\Delta} \mathrm{erfc}((\sigma^2/2-\Delta)/2\sigma^2)
\end{aligned}$$

where we have made the substitutions $\delta' = \delta - \Delta$ and $\delta'' = \delta' + \sigma^2$. This expression for $A(\Delta)$ will satisfy detailed balance, $A(\Delta) = e^{-\Delta} A(-\Delta)$.

In practice, both the energy difference and the error are being estimated from a finite set of data. Assume we have n estimates for the energy difference, $y_1, \ldots, y_n$. Estimates for the mean and variance are given by

$$\delta = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{5.6}$$

$$\chi^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i - \delta)^2 \tag{5.7}$$

and we have $\Delta = \langle \delta \rangle$ and $\sigma^2 = \langle \chi^2 \rangle$.

The average acceptance ratio can be written as integral over δ and χ^2. The probability distribution for the estimated error is a chi-squared distribution. An asymptotic solution can be formed by expanding $a(\delta, \chi^2)$ and performing the integrals to get the average acceptance ratio. This is set equal to a power series for $\exp(-\sigma^2/2)$, and by matching powers of σ we get the coefficients for the original series for $a(\delta, \chi^2)$. This series can by summed to get a Bessel function, hence we call it the Bessel acceptance formula. It is convenient to expand the log of the Bessel acceptance formula in powers of χ^2/n. The Bessel acceptance formula is then

$$a(\delta, \chi^2, n) = \min[1, \exp(-\delta - u_B)] \tag{5.8}$$

where

$$u_B = \frac{\chi^2}{2} + \frac{\chi^4}{4(n+1)} + \frac{\chi^6}{3(n+1)(n+3)} + \frac{\chi^8(5n+7)}{8(n+5)(n+3)(n+1)^2)} + \cdots \tag{5.9}$$

Note that as n gets large, only the first term is important, which is just the regular penalty method.

5.1.1 Other methods

There is another method for handling noise, originally proposed by Kennedy, Kuti, and Bhanot (**??**), that uses a power series expansion of $\exp[-\delta]$ to construct an unbiased acceptance ratio. It has an advantage over the penalty method in that it does not assume any particular distribution for the noise. The method has a major drawback in that it depends on the value of δ not becoming too large, and not just the error estimate for δ. This could severely restrict the maximum steps sizes for moving the nuclei in our simulations. Methods for dealing with this restriction has recently been addressed by **?**) and **?**), but we did not explore these extensions.

5.1.2 Handling noisy data

Using noisy data requires care in handling. Particularly, inappropriate reuse of any single estimated value can lead to biased results. For instance, in a classical simulation the energy difference would be computed, and one of the two energies involved would be used in accumulating the average energy. See the top of Figure 5.1 for an outline of such a simulation. However, this leads to a bias when noisy energies are involved. This can be seen by considering a negative fluctuation in the energy of the trial move. This will make the energy difference smaller (or more negative) and hence more likely to be accepted. Thus

the negative fluctuations would be preferentially added to the accumulated average, and bias the result downward.

This program outline could be corrected by computing a new value for the energy in the average. However, there is another arrangement that is more amenable to the energy difference methods of Chapter 3. The computation of the energy used in the average is the same quantity needed for the old energy in the next iteration. So that computation can be moved to the next iteration, as shown on the bottom in Figure 5.1.

Several points are illustrated with a system of two particle interacting via a Lennard-Jones potential with $\varepsilon = 0.1$ and $\sigma = 1.5$ [2]. The temperature of the system was 3160K ($\beta = 100$). Noise was simulated by adding a Gaussian random variable with known variance to every energy computation. Results for several algorithms versus noise level are shown in Figure 5.2. The top curve shows the bias that results from having no penalty method. The middle curve is the correct method, which we see is independent of the noise level. The bottom curve demonstrates the bias from reusing the energy involved in making the accept/reject decision.

The noise level of a system can be characterized by the relative noise parameter, $f = (\beta\sigma)^2 t/t_0$, where t is the computer time spent reducing the noise, and t_0 is the computer time spent on other pursuits, such as optimizing the VMC wavefunction or equilibrating the DMC runs. A small f means little time is being spent on reducing noise, where a large f means much time is being spent reducing noise. The efficiency of a CEIMC simulation can be written in terms of this parameter. Our paper gives an example of a double well potential, and finds the noise level that gives the maximum efficiency. Generally it falls around $\beta\sigma = 1-2$, with the optimial noise level increasing as the relative noise parameter increases. The one exception occurs when computing the first moment, which is sensitive to crossing the barrier between the double wells. These crossings are assisted by an increased noise level, hence the optimal noise level is much higher.

5.2 Pre-rejection

We can apply multi-level sampling ideas (see section 2.4.1 for an application to VMC) to our CEIMC simulations as well. The idea is to use an empirical potential to "pre-reject" moves that would cause particles to overlap and be rejected anyway.

In this case, the trial move is proposed and accepted/rejected based on a classical potential

$$A_1 = \min\left[1, \frac{T(R \to R')}{T(R' \to R)} \exp(-\beta \Delta V_{cl})\right] \tag{5.10}$$

[2] Here σ is the length scale for the LJ potential

Compute old Energy

loop over Classical steps

  loop over Number of molecules

    make trial move (translation and rotation of H_2 molecule)

    Compute trial Energy

    acceptance probability $= \min\left[1, \exp\left(-\beta\Delta E - (\beta\sigma)^2/2\right)\right]$

    accept/reject trial move

    if accept, set old Energy = trial Energy

    Use updated old Energy in average

  end loop

end loop

loop over Classical steps

  loop over Number of molecules

    make trial move (translation and rotation of H_2 molecule)

    Compute old Energy | Compute trial Energy

    acceptance probability $= \min\left[1, \exp\left(-\beta\Delta E - (\beta\sigma)^2/2\right)\right]$

    accept/reject trial move

    Use old Energy in average

  end loop

end loop

Figure 5.1: CEIMC program outlines. Boxes indicate quantum computations. The dashed box indicates a quantity saved from a previous computation. The top algorithm is incorrect. The bottom algorithm is correct.

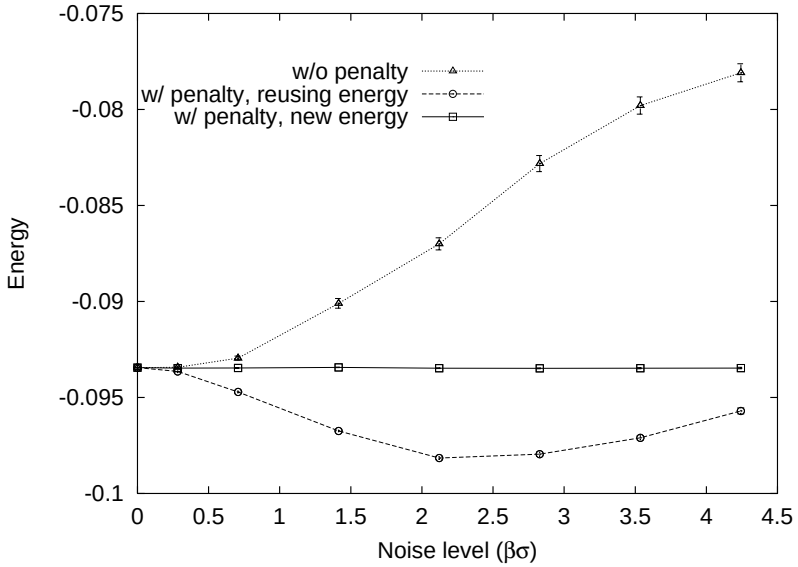


Figure 5.2: Examples on a Lennard-Jones potential with synthetic noise.

where $\Delta V_{cl} = V_{cl}(R') - V_{cl}(R)$ and T is the sampling probability for a move. If it is accepted at this first level, the QMC energy difference is computed and accepted with probability

$$A_2 = \min[1, \exp(-\beta \Delta V_{QMC} - u_B) \exp(\beta \Delta V_{cl})] \tag{5.11}$$

where u_B is noise penalty.

Compared to the cost of evaluating the QMC energy difference, computing the classical energy difference is free. Reducing the number of QMC energy difference evaluations is valuable in reducing the computer time required.

In Chapter 7, using the pre-rejection technique with a CEIMC-DMC simulation results in a first level (classical potential) acceptance ratio of 0.43, and a second level (quantum potential) acceptance ratio of 0.52. The penalty method rejects additional trial moves because of noise. If these rejections are counted as acceptances (ie, no penalty method or no noise), then the second level acceptance ratio would be 0.71. The classical potential is a fairly good representation for the DMC potential, and we can use that to reduce the number of DMC energy difference evaluations needed.

5.3 Trial Moves

Molecular moves are separated into translation, rotation and bond length changes.

The Silvera-Goldman potential is used as the empirical potential for pre-rejecting translational moves. Anisotropic potentials were tried for pre-rejecting rotational moves, but

Table 5.1: Efficiency of classical Monte Carlo for moving several particles at once. The table on the left is for low density system at $r_s = 3.0$ and T=5000K. The table on the right is for a high density system at $r_s = 1.8$ and T=3000K. The largest values of the efficiency are shown in boxes.

	Δ			
N_m	0.8	1.6	3.0	4.0
1	1.4	4.9	9.6	11.6
2	2.4	8.1	15.3	17.4
4	5.6	12.5	16.9	17.7
8	6.9	14.8	18.0	14.8
16	9.5	22.2	7.4	
32	11.9	14.8	2.7	

	Δ				
N_m	0.4	0.8	1.2	1.6	2.0
1	74	134	99	236	191
2	43	149	179	121	114
4	118	170	141	66	23
8	172	128	24		
16	155	52			
32	39				

they did not work very well. It is not clear whether this was from the the potentials being derived for isolated H_2-H_2 interaction, or from inaccuracy in the trial wave function.

Bond stretching moves were pre-rejected using the, essentially exact, H_2 intramolecular potential of **?**). The new bond length is sampled uniformly from a box of size Δ_b around the current position. Because of phase space factors we need to include a sampling probability of $T(R) = 1/R^2$ in the acceptance formula.

The trial move for classical Monte Carlo is usually presented as either moving one particle at a time, or all of the particles at once. However, we can move other numbers of particles as well. Table 5.1 shows the efficiency for a classical system with 32 H_2 molecules for two densities and temperatures. On the left is a low density system with $r_s = 3.0$ and at a temperature of 5000K, and on the right higher density system with $r_s = 1.8$ and a temperature of 3000K. For the lower density system, the highest efficiency occurs when moving half the molecules at a time. Relatively high efficiency can also be found moving 2, 4 or 8 at a time as well. For the higher density system, the most efficient regime shifts towards smaller step sizes and fewer number of particles moved at a time. [3]

5.4 Single H_2 molecule

The CEIMC method was applied to a single H_2 molecular in free space, at a temperature of 5000K. Exact results are obtained by integrating the potential of **?**). Results for the energy,

[3] These results are not generally applicable to classical MC simulations, since much more efficient implementations are possible for systems interacting with a two body potential.

Table 5.2: Results of CEIMC for isolated H_2 molecule at T=5000K.

	Energy	Virial	$\langle r \rangle$	$\sqrt{\langle r^2 \rangle}$
exact	-1.1630	0.0	1.57	1.60
VMC	-1.159(1)	-0.009(6)	1.56(2)	1.58(2)
DMC	-1.163(2)	-0.015(6)	1.58(2)	1.60(2)

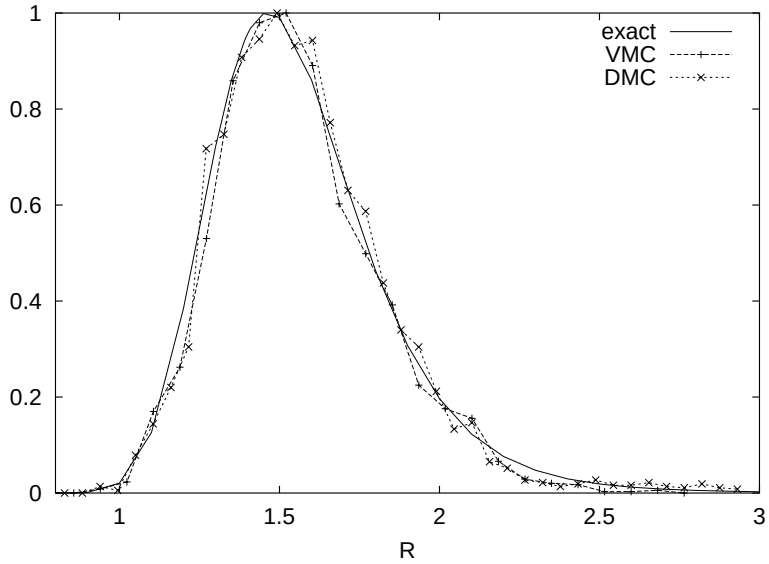


Figure 5.3: H_2 bond length distribution.

pressure, and first and second moments of the bond length are given in Table 5.2. The Virial column is computed by

$$\text{Virial} = [2\langle K \rangle + \langle V \rangle] \tag{5.12}$$

This is related by the virial theorem to the force on the nuclei, which should be zero for an isolated molecule. In Chapter 7, we will see this expression used to compute the pressure.

As we would expect, the VMC energy is higher than the exact energy. The other quantities are close to their expected values. Histograms of the bond length distribution are shown in Figure 5.3. Here again, both VMC and DMC reproduce the exact distribution well.

Chapter 6

Hard Spheres

A system of hard spheres is the simplest non-ideal many body system to study. Statistical Monte Carlo techniques and molecular dynamics were first applied to hard spheres (**??**). Additionally, some of the first applications of field theoretic methods to condensed matter systems were on hard spheres. More recently, the achievement of BEC in trapped atomic gases has renewed interest in the theory of the hard sphere Bose gas (**?**).

This chapter is a VMC and DMC study of a homogeneous, boson hard sphere fluid. Because they are bosons, there is no fixed node approximation in DMC, and the result can be made essentially exact. The approximations we must control are finite size effects and timestep error in DMC.

Applying perturbation theory to the Bose hard sphere gas yields the low density expansion,

$$E = 2\pi\rho[1 + C_1\sqrt{\rho} + C_2\rho\ln\rho + C_3\rho + ...] \tag{6.1}$$

where $C_1 = 128/(15\sqrt{\pi})$ and $C_2 = 8(4\pi/3 - \sqrt{3})$. Mean field theory gives the linear term. Straight forward application of perturbation theory yields the C_1 term, as was done by Huang, Yang, and Lee (**??**). The next higher order of perturbation theory diverges, but this was solved by including the depletion of the condensate by **?**) to get C_2. **?**) obtained the same results via a resummation technique. **?**) also obtained the logarithmic term. They also obtained the functional form for the series, which includes terms of the form $\rho^{n/2}$ and $\rho^{n/2}\log(\rho)$.

Renormalization group techniques have recently been applied to examine this divergence (**???**). In addition, Braaten and Nieto have calculated C_3 (**??**). This term is also first term that depends on more that the two body *s*-wave scattering length. It would require a solution to the three body scattering problem, which makes an explicit computation of that coupling constant difficult.

?) used VMC on an 864 particle system to find the fluid-solid transition density. **?**)

used the more accurate Green's Function Monte Carlo (GFMC) to calculate the energy of the solid and liquid phase near freezing to determine the freezing density. They used 256 particles. In the liquid state they computed four points ranging in density from 0.16 to 0.27.

Recently **?**) did DMC calculations on the homogeneous Bose gas, with various potentials, including the hard sphere potential. They used 500 particles and there was no mention of what DMC time step was used. These calculations were also used to make a fitting to the extended form of Eq. (6.1) (using terms up to $\rho^{5/2}$) by **?**).

There have been other attempts to get an equation of state by using various fitting techniques to combine the low density results and the GFMC results (**??**). As noted by **?**), the earlier work had an error and used only half the energy of the actual GFMC results. A new attempt at fitting the various functional forms was not done, but a new value for C_3 was estimated.

The Hamiltonian for this system is

$$H = -\frac{1}{2}\sum_i \nabla_i^2 + \sum_{i<j} v(|r_i - r_j|) \tag{6.2}$$

where

$$v(r) = \begin{cases} \infty & r < \sigma \\ 0 & r > \sigma \end{cases} \tag{6.3}$$

We have set $\hbar = m = \sigma = 1$ in all these calculations.

6.1 Wave function

An approximate wave function for the boson ground state that we use for a trial function is

$$\psi = \prod_{i<j} f(r_{ij}) \tag{6.4}$$

The individual functions are very similar to the ones used for hydrogen. The correlation function has a maximum range, r_{max}, beyond which f is constant. In order to be compatible with periodic boundary conditions, we require $r_{max} \le L/2$. The "cusp" condition is that the wave function must vanish linearly when two spheres get close, $f(r \to \sigma) \propto (r - \sigma)$. (Unlike the electronic case, the slope is not fixed.)

A change of variables will simplify these expressions. Let $x = r - \sigma$, $x_{max} = r_{max} - \sigma$ and $y = x/x_{max}$. Now y lies in the range $[0,1]$. In these variables, the boundary conditions on f are

$$\begin{aligned} f(y=0) &= 0 \\ f(y=1) &= 1 \end{aligned} \tag{6.5}$$

Table 6.1: Variational parameters for hard sphere gas

ρ	x_{max}	b_0	b_1	b_2	b_3	b_4
.2	1.92	-0.3209	0.25395	0.5624	0.0145	-.05123
.1	2.68	0.9173	1.995	0.8147	0.2269	0.0345
.05	2.68	-0.33	0.674	-0.12	0.056	0.0
.01	5.9152	-1.95	0.86267	-1.2982	-0.08135	-0.3152

$$f'(y=1) = f''(y=1) = 0 \tag{6.6}$$

$$\tag{6.7}$$

and the wave function is

$$f(y) = 3(y-y^2) + y^3 + y(y-1)^3 \sum_{i=0}^{4} b_i T_i(2y-1) \tag{6.8}$$

where T_i are Chebyshev polynomials, and b_i are variational parameters.

The parameters for each density are given in Table 6.1. They were obtained from optimization of the smallest system ($N = 40$). Then those same parameters were used for all system sizes at a particular density.

The cost of computing of the wave function and local energy is dominated by calculating the $N(N-1)/2$ interparticle distances. There are techniques for improving the scaling of computations of short range interactions to achieve $O(N)$. We used the cell method (**?**). The simulation box is divided into cubic cells and a list is made of all the particles in each cell. For simplicity, consider the case where each cell is larger than the cutoff distance, r_{max}. Then a particle in a cell will have a non-zero interaction only with the particles in the same cell and with particles in the neighboring cells. Particles in cells further away can be ignored. There is an overhead in computing and maintaining these lists. We used the cell method on systems with 500 particles and larger.

There is a deficiency with the trial wave function that leads to undersampling when three particles are in close proximity. In DMC, this leads to a large number of branching walkers to compensate for the undersampling, which invariably causes problems with maintaining a stable population. One solution is to use a guiding function, which differs from the trial wave function, for the diffusion and branching. Then a weight, ψ_T/ψ_G, is associated with each sample point. In this case the simplest guiding function is to use is $\psi_G = \psi_T^\alpha$. We found that $\alpha = 0.9$ was sufficient to make the population of walkers stable.

The DMC timestep errors should be local, and hence the same for all system sizes. We did timestep extrapolation on systems with $N = 40$ particles. The timestep errors were found to be linear in τ. The extrapolations to $\tau = 0$ are shown in Figure 6.1.

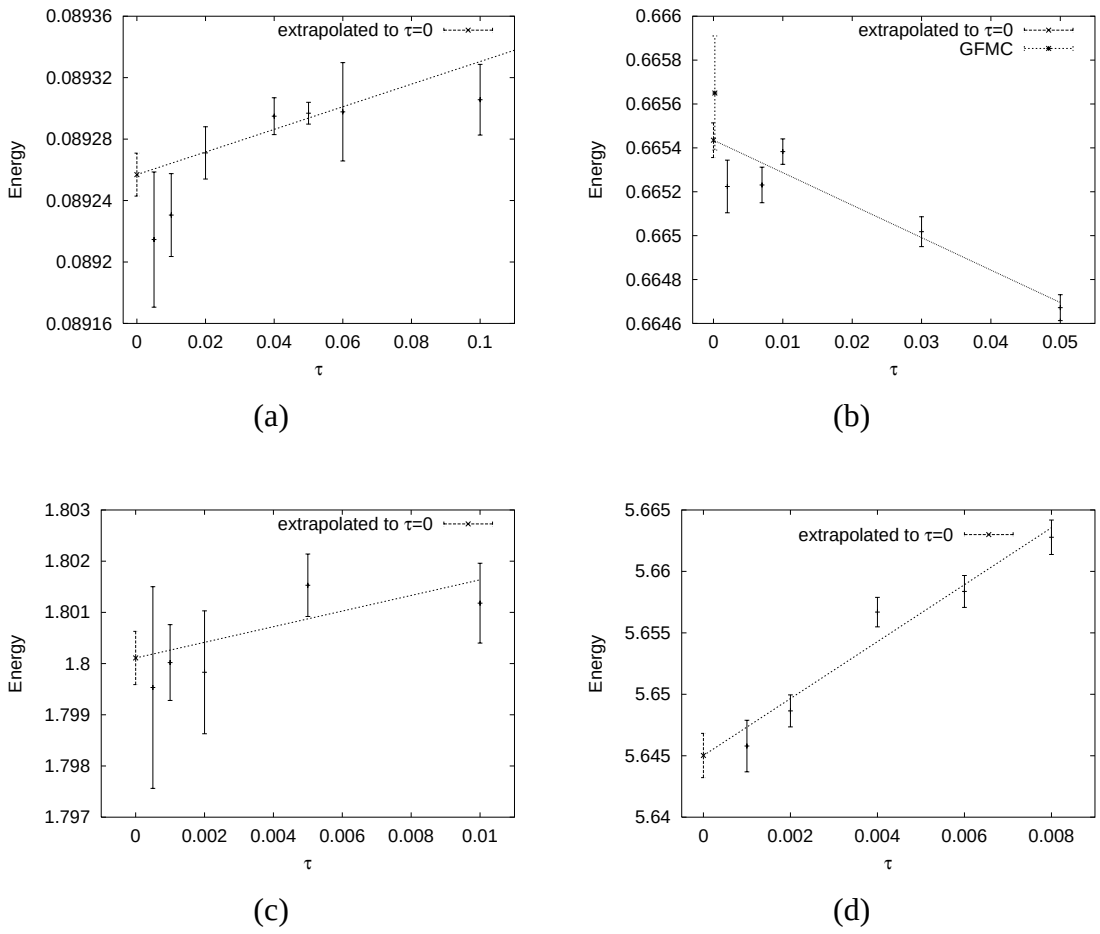


Figure 6.1: Time step error for (a) $\rho = 0.01$ (b) $\rho = 0.05$ (c) $\rho = 0.1$ (d) $\rho = 0.2$

Green's Function Monte Carlo (GFMC) uses a different decomposition of the Green's function that DMC. The principle advantage of GFMC is that is has no time step error. We ran GFMC at $\rho = .05$ and $N = 40$ to verify the timestep errors. The GFMC data point is shown at $\tau = 0$ in Figure 6.1. The importance sampling in GFMC is not as effective as in DMC, hence the variance is larger, and GFMC is less efficient than DMC. GFMC has an efficiency of about 7, whereas DMC has an efficiency of 240 at $\tau = .002$ and an efficiency of 570 at $\tau = .007$. Even with computing at several timesteps to extrapolate to zero, DMC is more efficient that GFMC.

6.2 Finite Size Effects

The main contribution to finite size effects in the energy is the long wavelength phonons. The functional form for their contribution depends on the small k behavior of the structure factor, $S(k)$. The energy can be written as

$$E = 4\pi \int_{k_b} k^2 dk \, \varepsilon(k) \tag{6.9}$$

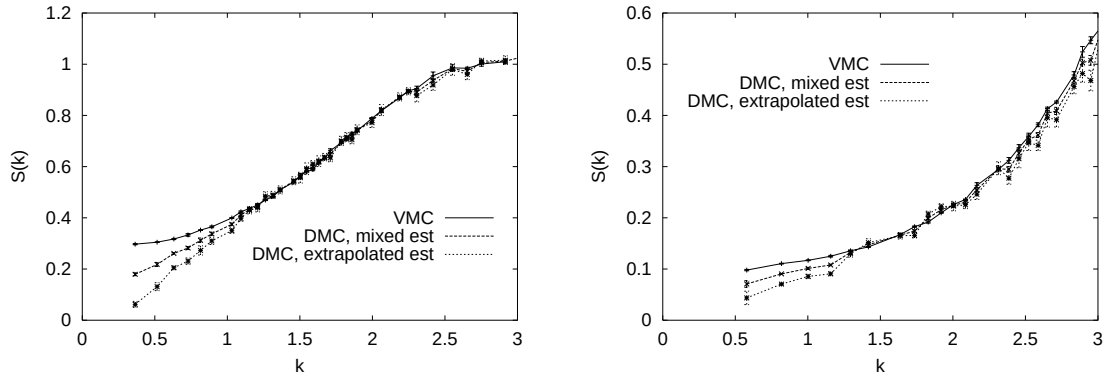
where $k_b = 2\pi/L$ is the small k cutoff due to the finite box size. The energy of the phonon excitations at small k is proportional to $S(k)$ (**?**). For a classical liquid, $S(k) \propto 1 + O(k^2)$. For a Bose fluid, $S(k)$ should be proportional to k and $S(k \to 0) = 0$.

The VMC wave function has no long range part, and so we expect it to behave like a classical fluid at small k. Integrating Eq (6.9), we get $E \propto k_b^3$, which is the same as scaling by $1/N$ for a fixed density. A more rigorous derivation of this scaling is given by **?**). The DMC algorithm should pick up the correct long wavelength behavior, leading to an $S(k)$ that is linear in k. Integrating Eq. (6.9) we get $E \propto k_b^4$. This then gives us $1/N^{4/3}$ scaling.

The small k behavior for S(k) can be seen nicely for $\rho = 0.05$, shown in Figure 6.2a. The graph shows that the VMC structure factor appears quadratic as expected. The DMC mixed estimator shows the $S(k)$ behaving linearly, but still not headed to zero. The extrapolated estimator looks like it over corrects and lowers S(k) too much.

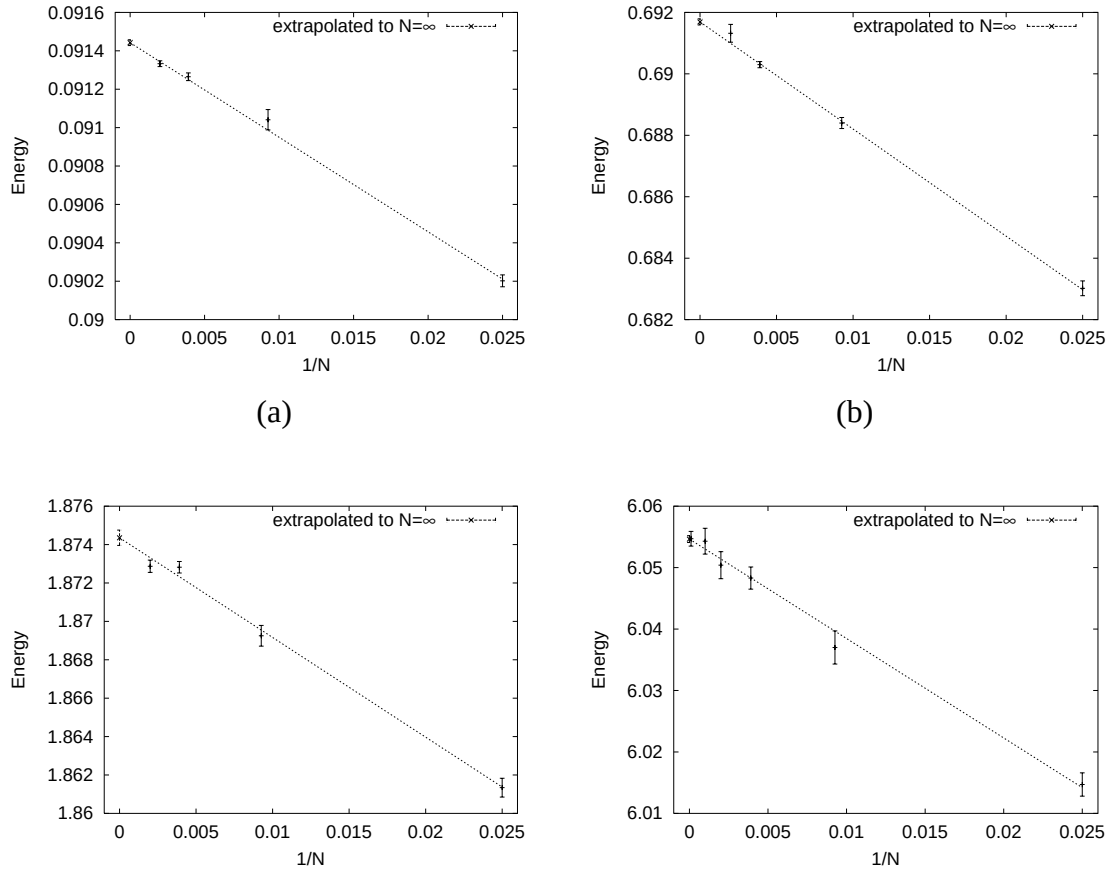
We did calculations for systems with 40, 108, 256, and 500 particles. For $\rho = .2$, additional VMC runs with $N = 10^3$ and $N = 10^4$ particles were done, and they are shown on the graph. The VMC energy nicely fits the $1/N$ behavior, as shown in Figure 6.3.

The DMC finite size extrapolation is shown in Figure 6.4. At higher densities, the DMC energy does not appear to have a $1/N^{4/3}$ dependence (or even $1/N$ dependence). However, the data is too sparse and noisy to make a good determination as to what the functional form should be, so we fit it to $1/N^{4/3}$. The final infinite system results are given in Table 6.2.



(a) (b)

Figure 6.2: S(k) for (a) $\rho = 0.05$ (b) $\rho = 0.2$



(a) (b)

(c) (d)

Figure 6.3: VMC finite size effects for (a) $\rho = 0.01$ (b) $\rho = 0.05$ (c) $\rho = 0.1$ (d) $\rho = 0.2$

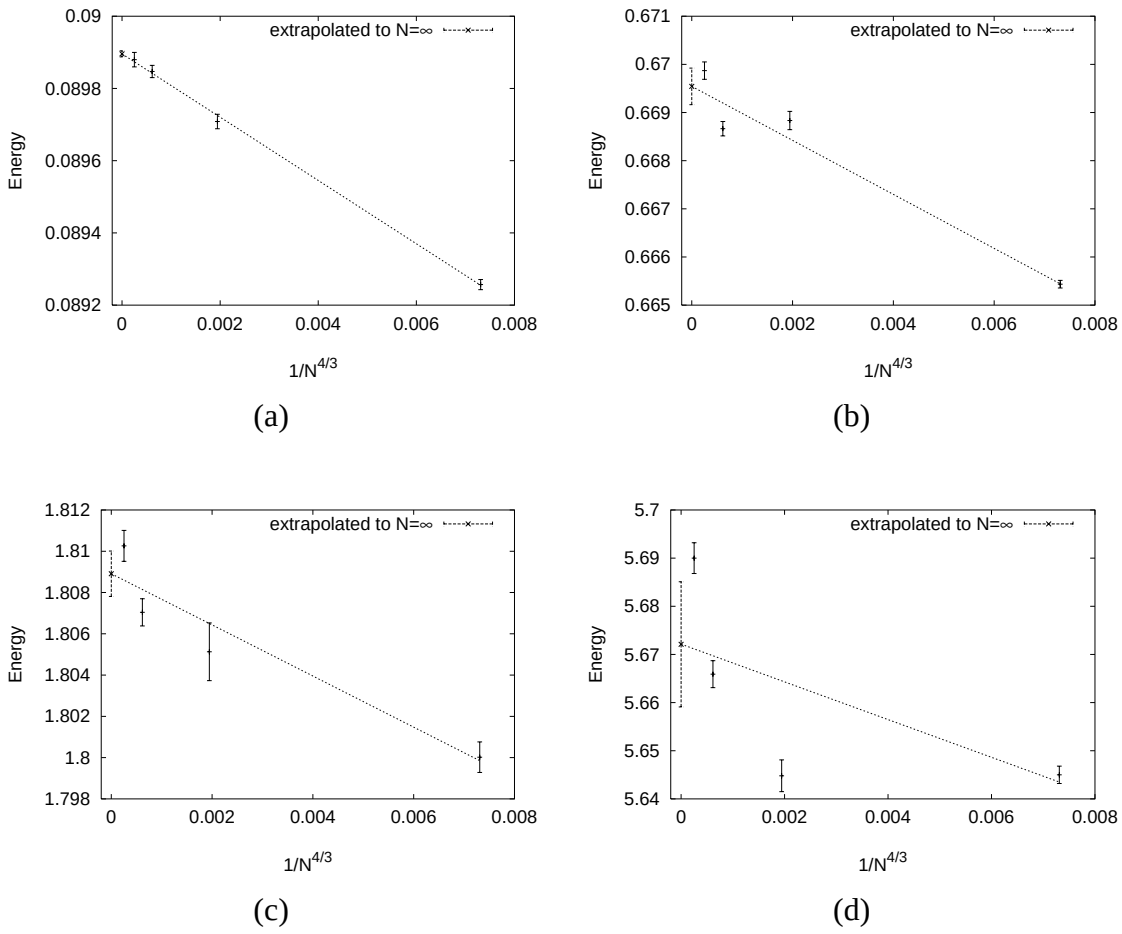


Figure 6.4: DMC finite size effects for (a) $\rho = 0.01$ (b) $\rho = 0.05$ (c) $\rho = 0.1$ (d) $\rho = 0.2$

Table 6.2: Energy extrapolated to infinite system size (in units of $\frac{\hbar^2}{m\sigma^2}$)

ρ	VMC	DMC	**?**)
.2	$6.0546(6)$	$5.67(1)$	
.1	$1.8744(4)$	$1.809(1)$	1.8130(35)
.05	$0.6917(1)$	$0.6690(4)$	0.6690(5)
.01	$9.144(2)\times10^{-2}$	$8.9896(8)\times10^{-2}$	$8.980(5)\times10^{-2}$

The long wavelength excitations will take a long time to sample, and their effect on the energy (and $S(k)$) may only be apparent with very long runs. And the time needed to sample them will increase with box size. The larger box sizes may be insufficiently converged, causing the energy to be too high. This may explain the apparent curvature in the wrong direction.

There are several approaches for resolving the problem at larger box sizes. The first is simply to perform even longer runs to see if the energy drops. Similarly, data for more system sizes would be helpful in outlining the functional form of the finite size dependence. Finally, explicit long range correlations could be added to the wave function, of the form proposed by **?**).

The energy versus density is shown in Figure 6.5, relative to the first order term , which is linear in the density. The low density exansion up to the C_1 term is also shown, as well as up to C_3, using the fitted value of 73.296 (**?**). It is clear from the graph, and was noted by **?**), that these additional terms by themselves do not help the expansion.

?) treated C_2 and C_3 as adjustable parameters and added two additional terms, $\rho^{5/2}\log(x)$ and $\rho^{5/2}$. They get a very good fit to the high density data, as seen in Figure 6.5.

The results of **?**) are also given in Table 6.2 for the three common densities computed. Their data and ours agree within the error bars.

6.3 Distribution Functions and Condensate Fraction

The two particle correlation function for all densities was calculated with a system size of 256 particles. The results for $g(r)$ using the extrapolated estimator are shown in Figure 6.6. We see the liquid shell structure developing as the density increases.

The single particle density matrix is the projection of a many-body wave function on to a single particle space. It is defined by

$$\rho_1(r,r') = \int dr_2 ... dr_N \; \psi(r,r_2,...,r_N)\psi^*(r',r_2,...,r_N) \tag{6.10}$$

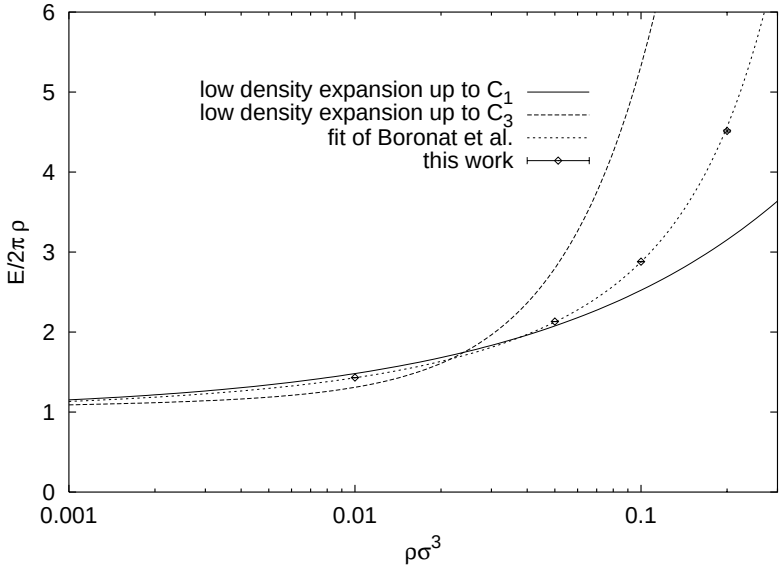


Figure 6.5: Energy vs. density

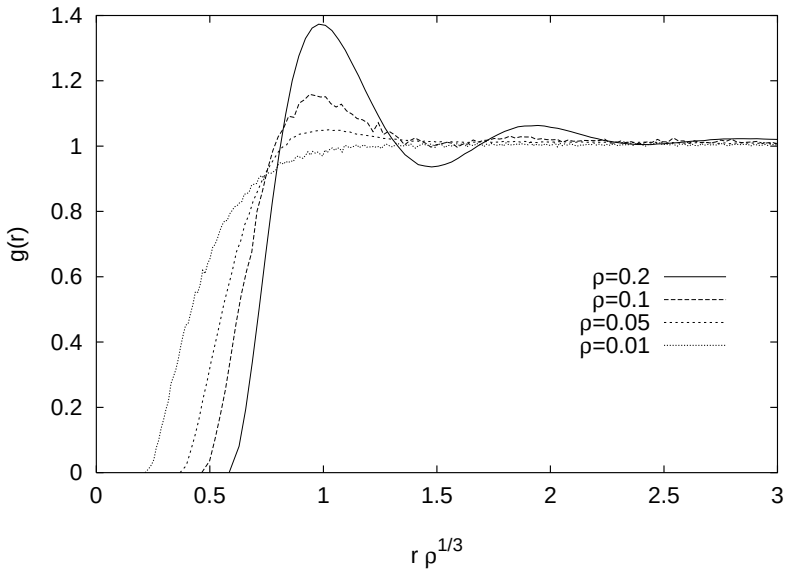


Figure 6.6: Pair distribution function for several densities

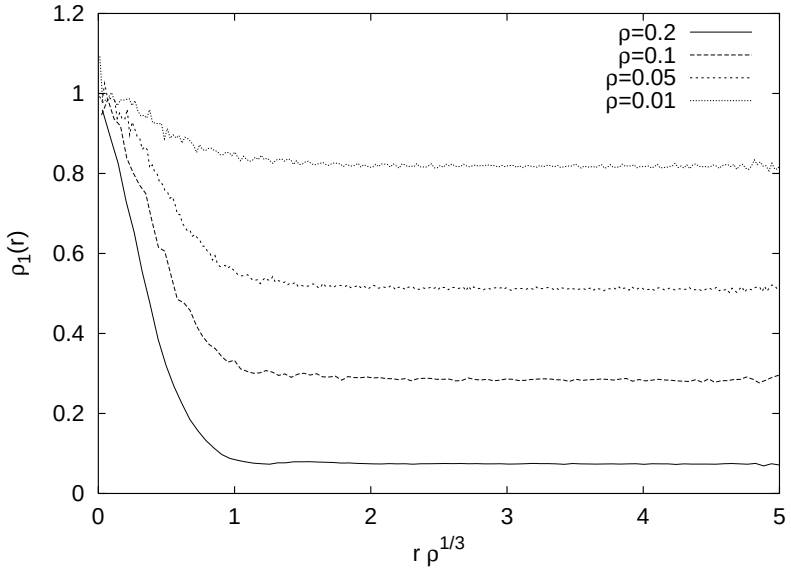


Figure 6.7: Single particle density matrix for several densities

In the homogeneous case, ρ_1 only depends on the distance between r and r'. The large r behavior of ρ_1 (or the $k = 0$ behavior of its Fourier transform, $n(k)$) is the condensate fraction. At zero temperature, the many body wave function is in the ground state. Because of interations, not all the particles are in the single body zero momentum state ($k = 0$ plane wave state in this case).

We used a method for sampling $\rho_1(r)$ which was given by **?**). The condensate fraction was obtained by integrating the single particle density matrix for distances greater than some cutoff, r_c, chosen to be where $\rho_1(r)$ had reached a plateau.

The condensate fraction is given in Table 6.3 and shown in Figure 6.8. Also shown in the figure is the GFMC result of $n_0 = 0.095(1)$ at $\rho = 0.2$. The low density expansion is given by

$$n_0 = 1 - \frac{8}{3\sqrt{\pi}}\rho^{1/2} \tag{6.11}$$

Similar to their treatment of the energy, **?**) added two additional terms, ρ and $\rho^{3/2}$. Their fit does a good job at higher densities, where, as expected, the low density expansion misses the full extent of the depletion.

Table 6.3: Condensate fraction

ρ	VMC	DMC (mixed)	Extrapolated
.2	0.1009(5)	0.0876(3)	0.0743(8)
.1	0.307(2)	0.2960(5)	0.285(2)
.05	0.563(2)	0.5401(4)	0.517(8)
.01	0.834(1)	0.826(1)	0.818(2)

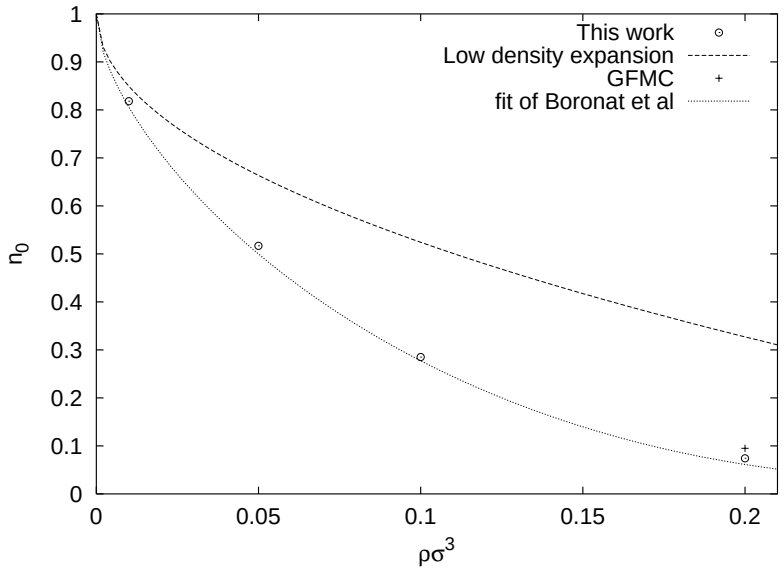


Figure 6.8: Condensate fraction vs. density

Chapter 7

Hydrogen

Hydrogen has been the subject of many experimental and theoretical studies. Theoretically, its simple electronic structure make it a favorable first target for various methods. Experimentally, hydrogen has been compressed by shock waves and also with a diamond anvil cell. We will present some CEIMC simulations and compare the results with those from one of the gas gun shock wave experiments.

7.1 Experiment

The high pressure experiments fall into two categories - transient compression from a shock wave or static compression from a diamond anvil cell. The shock wave experiments reach higher temperatures and pressures, but obtain more limited data. A high-velocity projectile hits a stationary target, inducing a shock wave in the target. The target is analyzed by the Hugoniot relations, derived by treating the shock wave as an ideal discontinuity and applying conservation of mass, momentum, and energy across it (**?**). The relations are then

$$P - P_0 = \rho_o u_s u_p \tag{7.1}$$

$$\rho = \rho_0 u_s/(u_s - u_p) \tag{7.2}$$

$$E - E_0 = \frac{1}{2}(V_0 - V)(P + P_0) \tag{7.3}$$

where E_0, P_0, V_0, and ρ_0 are the initial energy, pressure, volume, and density, respectively. The velocity of the shock wave is u_s and u_p is the velocity of the projectile driving the shock.

There are a number of methods for accelerating a projectile (**?**), but the two most prominent methods for hydrogen targets are the two stage light gas gun (**????**) and a large laser (**???**).

Recent advances make it possible to measure the temperature by light emission from the samples during compression (**?**). Measurement of the conductivity is also possible, used in recent experiments to detect metallic hydrogen (**?**).

The diamond anvil cell (DAC) is used to generate large static pressures. It has been used to study the fluid phase and several solid phases (**?**). It has also been used to determine the melting curve for hydrogen up to 500K (**??**).

7.2 Theory

Free energy models are typically based on the chemical picture, where molecules, atoms and various types of ions are all treated as different species of particles. Solving the physical picture, where the only fundamental particles are electrons and protons, is much more difficult (see Path Integral Monte Carlo below). The free energy of the various phases is constructed from a variety of fits to experimental data, equation of state data from reference systems (Lennard-Jones and hard sphere), and empirical and theoretical interaction potentials.

One of the best known models is that of Saumon and Chabrier (**??**). Extensive tables for astrophysical use were published by **?**). Another model was developed by **?**), to study the plasma (metal-insulator) transition.

The Path Integral Monte Carlo (PIMC) method is in principle the best method for simulations, since it treats both the electrons and protons quantum mechanically at non-zero temperature (**????**). The only major uncontrolled approximation is the location of the electron nodes, with problems and a solution similar to the fixed node method in DMC. **?**) have made progress in improving the nodal structure used in these calculations. PIMC is based on breaking up a thermal density matrix into a product of high temperature components, and consequently it works well at high temperature and becomes less efficient as the temperature decreases. About 5000K is currently the lower limit for PIMC calculations. Our CEIMC simulation technique should make a nice complement to the PIMC method.

There have also been path integral studies using empirical potentials, in order to examine the quantum effects of the nuclei on the system (**????**).

The Car-Parrinello method has been used to simulate this system (**????**). At low temperature, it it necessary to treat the nuclei with path integrals (**??**). Some studies used LDA with the Γ point approximation (using only one *k*-point for the integral over the Brillouin zone), which is not sufficient to converge the anisotropic behavior of the potential (**?**), and gives rise to unphysical planar structures (**?**).

7.3 Pressure and Kinetic Energy

The pressure is computed by a virial estimator based on the potential and kinetic energies

$$P = \frac{1}{3V}[2\langle K\rangle + \langle V\rangle] \tag{7.4}$$

where V is the volume and V is the potential energy. In these MC simulations, only the kinetic energy of the electrons is explicitly computed. The kinetic energy of the nuclei must also be added.

We are only considering hydrogen in the molecular state, and further assume that rotational and vibrational motion can be separated. The characteristic temperature for quantum effects for rotational motion is about 85 K for H_2 (**?**). At our simulation temperatures, we can use the classical expression for the rotational kinetic energy, $E_{\rm rot} = 2kT$.

The characteristic vibrational temperature for H_2 is $\theta_v = 6100$ K, so it is necessary to use the quantum expression for the vibrational kinetic energy. It is

$$E_{\rm vib} = \frac{\theta_v}{e^{-\theta_v/T} - 1} \tag{7.5}$$

For D_2, the characteristic temperature should be a factor of $\sqrt{2}$ lower.

Of course, these expressions are only valid for a free molecule. To truly treat the kinetic energy of the nuclei correctly in the interacting system, path integrals should be used for the nuclei.

7.4 Individual Configurations

We took several configurations from PIMC simulations at 5000K at two densities ($r_s = 1.86$ and $r_s = 2.0$), and compared the electronic energy using VMC, DMC, DFT-LDA, and some empirical potentials. The DFT-LDA results were obtained from a plane wave code using an energy cutoff of 60 Rydbergs, and using the Γ point approximation (**?**).

The empirical potentials are the Silvera-Goldman (**?**) and the Diep-Johnson (**??**). To these we added the energy from the Kolos (**?**) intramolecular potential to get the energy as a function of bond length variations. The Silvera-Goldman potential was obtained by fitting to low temperature experimental data, with pressures up to 20 kbar, and is isotropic. The Diep-Johnson potential is the most recent in a number of potentials for the isolated H_2-H_2 system. It was fit to the results of accurate quantum chemistry calculations for a number of H_2-H_2 configurations. It is an anisotropic potential.

The energies relative to an isolated H_2 molecule are shown in Figure 7.1. The first thing we notice is that the classical potentials are more accurate than VMC or DFT. The

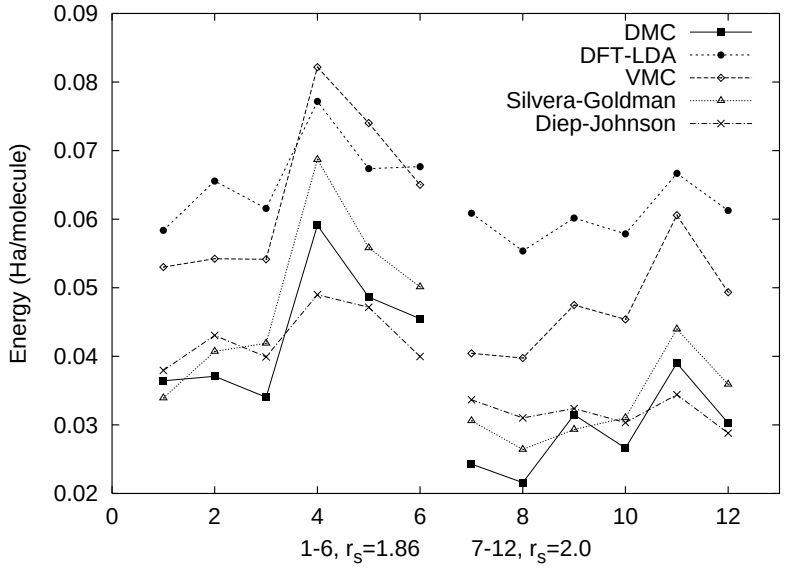


Figure 7.1: Electronic energy for several configurations computed by several methods. The energy is relative to an isolated H_2 molecule.

Silvera-Goldman mostly does a good job of reproducing the DMC results. [1] Some of the failures of the SG potential can be attributed to the lack of anisotropy. The isolated H_2-H_2 potential (Diep-Johnson) has much weaker interactions, compared with interactions in a denser system.

The PIMC method itself gives an average energy of about $0.07(3)$ Ha for both densities. Improvements in the fermion nodes appear to lower the energy (**??**), although the error bars are still quite large. There are also corrections to some internal approximations that lower the energy by an additional 0.02 Ha (**?**). These effects combined seem to bring the PIMC energy in rough agreement with the DMC energy.

We used the Silvera-Goldman potential for pre-rejection. As seen in the Figure 7.1, it resembles the DMC potential even though it lacks anisotropy. A hybrid potential was created by **?**), taking the isotropic part from a potential that was fit to high density, and combining that with the anisotropic part from one of the isolated H_2-H_2 potentials. We did not pursue this approach for constructing a better potential for pre-rejection.

7.5 Results

We obtained results from simulations at three state points, two of which can be compared with the gas gun data of **?**). The pressure is given in Table 7.1, with results from the gas gun

[1] It should be noted that we are taking the SG potential far from the temperature range it was fit to.

Table 7.1: Pressure from simulations and shock wave experiments

r_s	V(cc/mol)	T(K)	Pressure (Mbar)				
			Gasgun	S-C	S-G	CEIMC-VMC	CEIMC-DMC
2.100	6.92	4530	0.234	0.213	0.201	0.226(4)	0.225(3)
2.202	7.98	2820	0.120	0.125	0.116	0.105(6)	0.10(5)
1.800	4.36	3000	-	-	0.528	-	0.433(4)

experiments, the Saumon-Chabrier model, from simulations using the Silvera-Goldman potential, and from our CEIMC simulations. These state points are in the fluid molecular H_2 phase. For the gas gun experiments, the uncertainties in the mearsured temperatures are around 100-200K. The experimental uncertainties in the volume and pressure were not given, but previous work indicates that they are about 1-2% (**?**).

We did CEIMC calculations using VMC or DMC for computing the underlying electronic energy, which are the first such QMC calculations in this range. The simulations at $r_s = 2.1$ and $r_s = 1.8$ were done with 32 molecules, and the simulations at $r_s = 2.202$ were done with 16 molecules. We see that the pressures from VMC and DMC are very similar, and that for $r_s = 2.1$ we get good agreement with experiment.

There is a larger discrepancy with experiment at $r_s = 2.202$. The finite size effects are fairly large, especially with DMC. We also did simulations at $r_s = 2.1$ with 16 molecules and obtained pressures of $0.264(3)$ Mbar for CEIMC-VMC and $0.129(4)$ Mbar for CEIMC-DMC. The Silvera-Goldman potential showed much smaller finite size effects than the CEIMC simulations, so we that the electronic part of the simulation is largely responsible for the observed finite size effects.

The energies for all these systems are given in Table 7.2. The energy at $r_s = 2.1$ with 16 molecules for CEIMC-VMC is $0.0711(4)$ Ha and for CEIMC-DMC is $0.0721(8)$ Ha. The average molecular bond length is given in Table 7.3, and we see the bond length is compressed relative to the free molecule. The proton-proton distribution functions comparing CEIMC-VMC and CEIMC-DMC are shown in Figure 7.2. The VMC and DMC distribution functions look similar, with the first large intramolecular peak around $r = 1.4$ and the intermolecular peak around $r = 4.5$.

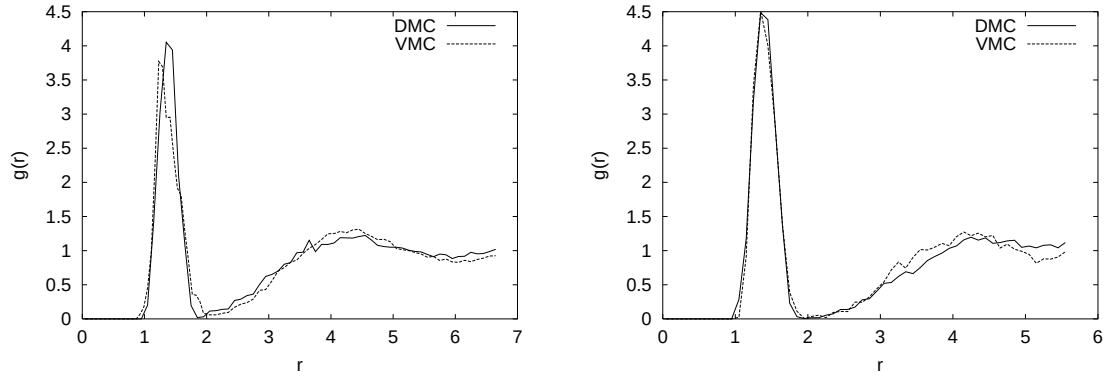
The CEIMC-VMC simulations at $r_s = 1.8$ and 3000 K never converged. Starting from a liquid state, the energy decreased the entire simulation. Looking at the configurations revealed they were forming a plane. It is not clear whether it was trying to freeze, or forming structures similar to those found in DFT-LDA calculations with insufficient Brillouin zone sampling (**??**). The CEIMC-DMC simulations did not appear to have any difficulty, so is seems the VMC behavior was due to inadequacies of the wave function.

Table 7.2: Energy from simulations and models, relative to the ground state of an isolated H_2 molecule. The H_2 column is a single thermally excited molecule plus the quantum vibrational KE.

r_s	V(cc/mol)	T(K)	Energy (Ha/molecule)				
			H_2	S-C	S-G	CEIMC-VMC	CEIMC-DMC
2.100	6.92	4530	0.0493	0.0643	0.0689	0.0663(8)	0.0617(2)
2.202	7.98	2820	0.0290	0.0367	0.0408	0.0305(8)	0.0334(9)
1.800	4.36	3000	0.0311	-	0.0722	-	0.055(1)

Table 7.3: Average molecular H_2 bond length. The H_2 column is a single thermally excited molecule in free space.

r_s	T(K)	Average bond length (Bohr)		
		H_2	CEIMC-VMC	CEIMC-DMC
2.100	4530	1.550	1.431(1)	1.413(3)
2.202	2820	1.486	1.443(1)	1.429(6)
1.800	3000	1.492	-	1.410(1)



(a) (b)

Figure 7.2: Proton pair distribution function $g(r)$ for (a) $r_s = 2.1$ and T=4530 K (b) $r_s = 2.202$ and T=2820 K

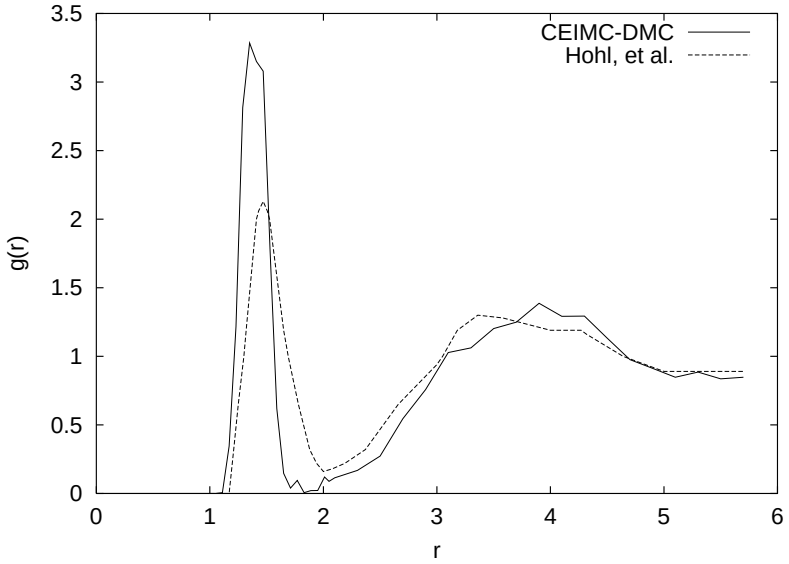


Figure 7.3: The proton pair distribution function, $g(r)$, close to $r_s = 1.8$ and T = 3000K.

?) did DFT-LDA simulations at $r_s = 1.78$ and T=3000K, which is very close to our simulations at $r_s = 1.8$. The resulting proton-proton distribution functions are compared in Figure 7.3. The discrepancy between CEIMC and LDA in the intramolecular portion of the curve has several possible causes. On the CEIMC side, it may be due to an insufficiently long run or due to the molecular nature of the wave function, which does not allow dissociation. The deficiencies of LDA may account for it preferring fewer and less tightly bound molecules. LDA is known to overestimate the bond length of a free hydrogen molecule (**?**), which would account for the shifted location of the bond length peak.

7.6 Simulation analysis

We also recorded some diagnostic information about the workings of the simulation, such as the average noise level, the relative noise parameter $f = (\beta\sigma)^2 t/t_0$, and a quantity called the additional noise rejection ratio, η. When a move is rejected with the penalty method, it is useful to recompute the acceptance decision with the same random number and without the noise penalty. If the move would have been accepted without the noise penalty, it is considered a rejection due to noise (as opposed to a rejection due to a trial move that increases the energy). This can be used to monitor the effects of the noise on the simulation. The additional noise rejection ratio is defined as

$$\eta = \frac{N_{\text{noise\_rej}}}{N_{\text{noise\_rej}} + N_{\text{accept}}} \tag{7.6}$$

Table 7.4: Simulation quantities ordered according to average noise level, $\beta\sigma$. The time column is the time for a single quantum step in minutes on an SGI Origin 2000. N is the number of molecules in the simulation.

r_s	T(K)	N	QMC	$\beta\sigma$	f	η	σ_d^2/σ_{ts}^2	time (min)
2.100	4530	16	VMC	0.68	0.17	0.11	2.2	18
2.202	2820	16	VMC	0.70	0.27	0.13	3.2	21
2.100	4530	32	VMC	0.90	0.29	0.16	2.3	70
1.800	3000	32	VMC	0.91	0.30	0.15	7.7	89
2.100	4530	16	DMC	1.62	2.28	0.28	-	76
2.100	4530	32	DMC	1.74	5.30	0.29	-	440
2.202	2820	16	DMC	2.02	5.33	0.40	-	92
1.800	3000	32	DMC	2.42	13.1	0.42	-	510

If η is small, the noise is causing few additional rejections. If η is 1/2, the noise is causing as many moves to be rejected as accepted. As $\eta \to 1$, the noise is causing many moves to be rejected.

Table 7.4 show the noise level $(\beta\sigma)$, the relative noise parameter, f, the additional noise rejection ratio, a ratio of the error level for the direct method and the two-sided method, and the time for a single quantum step. Looking at f, we see it is small for VMC and large for DMC. This is because of VMC optimization takes a proportionately larger amount of time in the VMC run than in the DMC run.

We used the two-sided method for computing energy differences of trial moves with VMC, but only used the direct method with DMC. The column headed σ_d^2/σ_{ts}^2 shows how the efficiency of computing the energy difference is improved by using the two-sided method. This improvement is only in the energy difference part of the total time, the optimization time is unaffected (and is a large part of the run time, since the f parameter is small). In Chapter 3, there was an example showing that the two-sided method was not as effective for DMC as for VMC. But in these simulations the DMC runs have a much larger f parameter, so even small reductions in the noise level would have an impact on the run time.

Some of DMC energy differences had values of noise many times greater than the average, which may be due to an instability in the DMC algorithm. We removed these outliers in computing the average noise level.

We tried the method for even lower temperatures with a simulation at T=800 K and $r_s = 1.8$ and it had a promising start, but after a while the acceptance ratio dropped and we were unable to get any usable data.

7.7 Future Work

The finite size effects in DMC need to be resolved. Using Ewald sums for computing the Coulomb interaction might help alleviate some of the finite size effects. Extending the wave function to allow for dissociated molecules and to provide for ionization would help make the simulations more accurate, particularly at higher temperatures and pressures.

Chapter 8

Conclusions

In this work we have developed new methods for increasing the scope of QMC calculations, and for increasing their efficiency. Variational Monte Carlo depends on optimizing parameters, but the presence of noise makes it difficult. We have examined several different kinds of optimization approaches and compared them. Further work should improve these methods even more.

The boson hard sphere model is an important theoretical model. We have performed "computational experiments" to obtain the ground state energy of this model. The effects of long range correlation on the energy are masked by the current uncertainty in the infinite system size results. However, if more accurate results are desired, the nature of the long range correlations and their effect on the energy will need to be more clearly resolved.

As a method for including increasingly more detailed and accurate physical effects in our simulations, we have developed the Coupled Electronic-Ionic Monte Carlo method. The central idea is simple, but several supporting developments were needed to make it computationally feasible. The penalty method enables use of energy differences with a noise level of approximately $k_B T$, rather than needing noise smaller than some fraction of $k_B T$ to avoid bias. The two sided energy difference method can stably compute these energy differences.

The CEIMC method was applied to a system of molecular hydrogen at a few state points. It shows promise for generating accurate simulation results.

Appendix A

Determinant Properties

The elements of the Slater matrix are $D_{ij} = \phi_j(r_i)$. The Slater determinant looks like

$$\begin{vmatrix} \phi_1(r_1) & \dots & \phi_n(r_1) \\ \vdots & \ddots & \vdots \\ \phi_1(r_n) & \dots & \phi_n(r_n) \end{vmatrix} \tag{A.1}$$

We assume that the single particle orbitals depend only on a single coordinate (ie, no back-flow).

The determinant of a matrix can be computed using the expansion by cofactors. This expands the determinant of an $N \times N$ matrix into a sum of N determinants of $(N-1) \times (N-1)$ matrices. As a recursive algorithm for computing the determinant, it is not very efficient, but for theoretical analysis, it is very useful for isolating the influence of a single row or column.

Define the cofactors of a matrix M to be

$$c_{ij} = (-1)^{i+j} \left| M_{ij} \right| \tag{A.2}$$

where the matrix formed by c_{ij} is called the cofactor matrix. The matrix M_{ij} is an $(N-1) \times (N-1)$ matrix formed by removing row i and column j from A. The determinant of A can then be written as

$$|A| = \sum_j a_{kj} c_{kj} = \sum_i a_{ik} c_{ik} \tag{A.3}$$

for $k = 1 \dots N$. The transpose of the cofactor matrix is called the adjoint of A. Now the adjoint is related to the inverse by

$$\text{adj } A = |A| A^{-1} \tag{A.4}$$

To compute the ratio of determinants, expand the determinant of $D(r_k')$ in cofactors about the kth row. Note that then the cofactors have no dependence on r_k'.

$$
\begin{aligned}
\left|D(r'_k)\right| &= \sum_i \phi_i(r'_k) c_{ki} \\
&= \sum_i \phi_i(r'_k) \left|D(r_k)\right| (D^{-1}(r_k))_{ik} \\
\frac{\left|D(r'_k)\right|}{|D(r_k)|} &= \sum_i \phi_i(r'_k) (D^{-1}(r_k))_{ik}
\end{aligned}
$$

If a move is accepted, the inverse matrix can be updated in $O(N^2)$ time (rather than $O(N^3)$ for recomputing the inverse). The formula for updating an inverse if only a single row (or column) changes was given by **?**). Let q be the ratio of determinants given above. Row k merely needs to be updated to reflect the new determinant, $D_{kj}^{-1} = D_{kj}^{-1}/q$. The other rows are updated as

$$
D_{ij}^{-1} = D_{ij}^{-1} - \frac{D_{ik}^{-1}}{q} \sum_l D_{lj}^{-1} \phi_l(r'_k) \quad i \neq k \tag{A.5}
$$

Appendix B

Elements of the Local Energy

The wave function has the form

$$\psi_T = D\exp\left[-U\right] \tag{B.1}$$

where D is the product of a spin up and a spin down Slater determinant and

$$U = \sum_{i<j} u(r_{ij}). \tag{B.2}$$

The local energy is then

$$E_{\rm L} = \frac{1}{2}\nabla^2 U - \frac{1}{2}\nabla U\cdot\nabla U - \frac{1}{2}\left(\frac{\nabla^2 D}{D}\right) + \left(\frac{\nabla D}{D}\right)\cdot\nabla U + V \tag{B.3}$$

In Diffusion Monte Carlo, we need the quantum force, $F_{\rm Q} = \nabla \ln|\psi|^2$.

$$F_{\rm Q} = 2\left(\frac{\nabla D}{D}\right) - 2\nabla U \tag{B.4}$$

The derivatives of the Jastrow factors are

$$\nabla_k U_{ee} = \sum_{i\neq k} u'_{ee}(r_{ik})\frac{\mathbf{r_i}-\mathbf{r_k}}{r_{ik}} \tag{B.5}$$

$$\nabla_k U_{ne} = \sum_{\alpha=1}^{M} u'_{ne}(r_{k\alpha})\frac{\mathbf{r_k}-\mathbf{R}_\alpha}{r_{k\alpha}} \tag{B.6}$$

$$\nabla_k^2 U_{ee} = \sum_{i\neq k}\frac{2}{r_{ik}}u'_{ee}(r_{ik}) + u''_{ee}(r_{ik}) \tag{B.7}$$

$$\nabla_k^2 U_{ne} = \sum_{\alpha=1}^{M}\frac{2}{r_{k\alpha}}u'_{ne}(r_{k\alpha}) + u''_{ne}(r_{k\alpha}) \tag{B.8}$$

For the gradient with respect to particle k, expand the determinant by cofactors about row k. Then the cofactors have no r_k dependence.

$$\frac{\nabla_k|d|}{|d|} = \sum_i \left[\nabla_k\phi_i(r_k)\right] d_{ik}^{-1} \tag{B.9}$$

$$\frac{\nabla_k^2 |d|}{|d|} = \sum_i \left[\nabla_k^2 \phi_i(r_k)\right] d_{ik}^{-1} \tag{B.10}$$

Appendix C

Cusp Condition

When two Coulomb particles get close, the potential has a $1/r$ singularity. The wave function must have the correct form to cancel this singularity. First, consider an electron and a nucleus. The relevant part of the Schrödinger equation is

$$\left[-\frac{1}{2M}\nabla_n^2 - \frac{1}{2}\nabla_e^2 - \frac{Ze^2}{r}\right]\psi = E\psi \tag{C.1}$$

where M is the nuclear mass and Z is the nuclear charge. Assume that $M \gg m_e$, so the first term can be ignored. Write the second term in spherical coordinates and we get

$$-\frac{1}{2}\psi'' - \frac{1}{r}\left(Ze^2\psi + \psi'\right) = E\psi \tag{C.2}$$

In order for the singularity to cancel at small r, the term multiplying $1/r$ must vanish. So we have

$$\frac{1}{\psi}\psi' = -Ze^2 \tag{C.3}$$

If $\psi = e^{-cr}$ we must have $c = Ze^2$.

For the case of two electrons, the Schrödinger equation is

$$\left[-\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 + \frac{e^2}{r_{12}}\right]\psi = E\psi \tag{C.4}$$

Switching to relative coordinates $r_{12} = r_1 - r_2$ gives us

$$\left[-\nabla_{12}^2 + \frac{e^2}{r_{12}}\right]\psi = E\psi \tag{C.5}$$

Electrons with unlike spins (no antisymmetry requirement) have an extra factor of $1/2$ in the cusp condition compared with the electron-nucleus case. So we have $c = -e^2/2$.

In the antisymmetric case, the electrons will be in a relative p state, reducing the cusp condition by $1/2$, so $c = -e^2/4$. Having the correct cusp for like spin electrons gains very little in the energy or the variance, since the antisymmetry requirement keeps them apart anyway.

Vita

Mark Dewing was born on May 18, 1971 in San Diego, California. He received a B. S. in physics from Michigan Technological University in 1993. He received an M. S. in physics from the University of Illinois at Urbana-Champaign in 1995.